

**PCT**  
WELTORGANISATION FÜR GEISTIGES EIGENTUM  
Internationales Büro  
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)



<p>(51) Internationale Patentklassifikation <sup>6</sup> : <b>C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027</b></p>	<b>A2</b>	<p>(11) Internationale Veröffentlichungsnummer: <b>WO 99/33998</b></p> <p>(43) Internationales Veröffentlichungsdatum: <b>8. Juli 1999 (08.07.99)</b></p>		
<table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top; padding: 5px;"> <p>(21) Internationales Aktenzeichen: <b>PCT/EP98/08216</b></p> <p>(22) Internationales Anmeldedatum: <b>22. Dezember 1998 (22.12.98)</b></p> <p>(30) Prioritätsdaten: <b>197 57 984.1      24. Dezember 1997 (24.12.97)    DE</b></p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): <b>BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).</b></p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): <b>HAGEN, Gustav [DE/DE]; Bertha-von-Sutner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggen-dorfstrasse 59, D-51061 Köln (DE).</b></p> <p>(74) Gemeinsamer Vertreter: <b>BAYER      AKTIENGE-SELLSCHAFT; D-51368 Leverkusen (DE).</b></p> </td> <td style="width: 50%; vertical-align: top; padding: 5px;"> <p>(81) Bestimmungsstaaten: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Veröffentlicht</b> <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p> </td> </tr> </table>			<p>(21) Internationales Aktenzeichen: <b>PCT/EP98/08216</b></p> <p>(22) Internationales Anmeldedatum: <b>22. Dezember 1998 (22.12.98)</b></p> <p>(30) Prioritätsdaten: <b>197 57 984.1      24. Dezember 1997 (24.12.97)    DE</b></p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): <b>BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).</b></p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): <b>HAGEN, Gustav [DE/DE]; Bertha-von-Sutner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggen-dorfstrasse 59, D-51061 Köln (DE).</b></p> <p>(74) Gemeinsamer Vertreter: <b>BAYER      AKTIENGE-SELLSCHAFT; D-51368 Leverkusen (DE).</b></p>	<p>(81) Bestimmungsstaaten: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Veröffentlicht</b> <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>
<p>(21) Internationales Aktenzeichen: <b>PCT/EP98/08216</b></p> <p>(22) Internationales Anmeldedatum: <b>22. Dezember 1998 (22.12.98)</b></p> <p>(30) Prioritätsdaten: <b>197 57 984.1      24. Dezember 1997 (24.12.97)    DE</b></p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): <b>BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).</b></p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): <b>HAGEN, Gustav [DE/DE]; Bertha-von-Sutner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggen-dorfstrasse 59, D-51061 Köln (DE).</b></p> <p>(74) Gemeinsamer Vertreter: <b>BAYER      AKTIENGE-SELLSCHAFT; D-51368 Leverkusen (DE).</b></p>	<p>(81) Bestimmungsstaaten: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Veröffentlicht</b> <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>			
<p>(54) Title: <b>REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF</b></p> <p>(54) Bezeichnung: <b>REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG</b></p>				
<p style="text-align: center;">A. FULL LENGTH</p>				
<p>(57) Abstract</p> <p>The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.</p>				

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

**LEDIGLICH ZUR INFORMATION**

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauritanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen  
Telomerase-Untereinheit und deren diagnostische und therapeutische Verwen-  
dung

5     Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten  
10     Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15     Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20     Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

25     Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der  
30     Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

10

15 Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

#### Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzelern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

25

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

30

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

#### 10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34<sup>+</sup>38<sup>+</sup>-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischen Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

5      Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase  
10      enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus  
15      von Tumorzellen zur Stabilisation ihrer Telomere.

Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzweiligen Strahlen schädigen nicht nur die Tumorzellen,  
20      sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle  
25      Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden  
30      dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapie von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

#### Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und posttranslationale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1“) spezifisch gebunden werden.

25

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements“) be-

30

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebs- oder zellspezifische Promotoreigenschaften.

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).



Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon  
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere  
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).  
20

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- $\alpha$ , IFN- $\beta$ , IFN- $\gamma$ , TNF, TNF- $\alpha$ , TNF- $\beta$ , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls  
30 zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 $\alpha$ , - $\beta$ ), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle  $\beta$ -Glukuronidase, pflanzliche  $\beta$ -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle  $\beta$ -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC),  $\beta$ -Galaktosidase ( $\beta$ -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH),  $\beta$ -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch  
5 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch  
10 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen,  
15 sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.  
20

Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.  
25

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

30 Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilsequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF  $\beta$ , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

20

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

30 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen  
5 Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen,  
15 die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der  
20 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

25 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

30

- A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten;
- C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Sprague Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten  $\lambda$ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene  $\lambda$ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des  
10 rekombinanten Phagens. Die Pfeile kennzeichnen die zwei  $\lambda$ -Arme des Vektors EMBL3 Sp6/T7.

Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten  $\lambda$ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

15 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene  $\lambda$ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur  
20 14: nicht verdaut.

25 B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

10

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5'-GTTAAGTTGTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

15

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

20

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25

Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymchnittstellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

30



die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

10

15

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

20

25

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgespligte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

30

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite ist die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

25 Tab. 1: Exon Intron Übergänge des hTC-Gens

Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind  
30 gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in

10 Intron 2.

Tab. 1

3' Acceptor Sequence			5' Donor Sequence					
Intron	Exon	Exon No.	bp	Intron	Intron No.	bp	Intr on	bp
5' flanking Region								
caggggcctcccccgtag	GTTCAGGCAGCGCTGCGT	1	281	CGCCCCCTCCTTCGCCAG	1	104		
catgtcctctctgctttag	GTGCTCTGCTGAAGGAGC	2	1354	TGGCTGCGCAGGAGCCAG	2	8616		
gaggggctctctattgtag	GGGTTGGCTGTGTTCCGGC	3	196	TGCAAGCATTGGAATCAG	3	2089		
cccatgtgtccccgcgtag	ACAGCACTTGAAGAGGCTG	4	181	GTTCCGCAGAGAAAAGAGG	4	687		
ctcgcctccactcacacag	GCCGAGGCTCTCACCTCGA	5	180	TGAGCTGTACTTTTGTCAAG	5	494		
ccctctcctctgccccgtag	GTGDAATGTGACGGGCGCGT	6	156	CAAGGCCTTCAAGAGCCAC	6	>4660		18
ctccccgtctgttttcgtag	GTCTCTACCTTGACAGACC	7	96	TGCCGTGTCATCGAGCAG	7	980		
ctggtgtcttccccgtag	AGCTCTCCCTGATGAGG	8	86	CCGTGCGCATCAGGGGCAA	8	2484		
gtattttcccttatttttag	GTCTACGTCCAGTGCCAG	9	114	CGGGGATTCGGCGGACGG	9	1984		
cattgccccctctgcttag	GCTGCTCCTGCGTTTGGTG	10	72	ACGGAAACCTTCCTCAG	10	1871		
attccccctgtgtcttag	GACCCTGGTCCGAGGTGC	11	189	TGCAGACGACTACTCCAG	11	3804		
tcctttctggcgactcttag	CTATGCCGGACCTCCATC	12	127	CCTGTTTCTGGATTTCAG	12	880		
ctgtccgcacatctcttag	GTGAACAGCCCTCCAGACGG	13	62	TCCTGCTGCAGGCGTACAG	13	3187		
agcctctgttttccccgtag	GTTCACGCATGTGTGCTG	14	125	CTGAAAGCCAGAACGCGAG	14	781		
tcgtattttggcccccgtag	GGATGCTGCTGGGGCCAA	15	138	CTGGGCTCACTCAGGACAG	15	536		
	CCCAGACGCAGCTGAGTCG	16	664	TTTTTCAGTTTGGAAAAA	3' flanking Region			

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Sp1	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
API	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

### Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

#### Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschrte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

## 5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca  $1,5 \times 10^6$  Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa  
10 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml  
15 Heringssperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit  $1,5 \times 10^6$  cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte  
20 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.*  
25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen  
30 Experimenten jeweils 1 bis  $1,5 \times 10^6$  Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach  $\lambda$ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umklontiert (Beispiel 4).

### Beispiel 3

Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde  $\lambda$ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).

Da die isolierte  $\lambda$ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

### Beispiel 4

Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einem das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.



- Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umklont. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.
- Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subklont. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.
- Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subklont.
- Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subklont. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10

15

Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-ACTATAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen

20

25

30

kloniert.

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTTG-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATAACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten

5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking

10 Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen genomischen hTC-Region identifiziert (s. Fig. 7).

15 Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten

20 Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch

25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

## Contig1:

5 ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCCAC ACGCCAGCCT TGGTGACAGA 70  
 ATGAGACCCCT GTCTCAAAAA AAAAAAATAA AATTGAAATA ATATAAGCA TCTTCTCTGG CCACAGTGA 140  
 ACAAAAACCG AAATCAACAA CAAGAGGAAT TTTGAAAAC ATACAAACAC ATGAAAAATA AACATATAC 210  
 TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTAAAAAGGA AATTGAAAAA TTTATTTAAG CAAATGATA 280  
 CGGAAACATA ACCTCTCAAA ACCCAGCGTA TACAGCAAAA GCAGTGCTAA GAAGGAAATT TATAGCTATA 350  
 AGCAGCTACA TCAAAAAAGT AGAAAAAGCA GGCAGCTGG CTCATGCCGT TAATCCCAGC ACTTTGGGAG 420  
 10 GCCAAGGCGG GCAGATCGCC TGAGGTGAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTG 490  
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG 560  
 GCAGGATAAC CGCTTGAAAC CAGGAGGTGG AGGTTCGGT GAGCCGGGAT TGGCCCATTT GACTCCAGCC 630  
 TGGGTAAACA GAGTGAAACC CTGTCTCAAG AAAAAAATAA AAGTAGAAAA ACTTAAAAAT ACACCTTAAT 700  
 GATGCACCTT AAAGAACTAG AAAAGCAAGA GCAAACTAAA CCTAAAATTG GTAAAAAGAA AGAAATAATA 770  
 15 AAGATCAGAG CAGAAATAAA TGAAACTGAA AGATAACAAT ACAAAAGATC AACAAAATTA AAAGTTGGTT 840  
 TTTTGAAAAG ATAAACAAAA TTGACAAACC TTTGCCAGA CTAAGAAAAA AGGAAAGAAG ACCTAAATAA 910  
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CACTGATAC CACAGAAATT CAAAGGATCA CTAGAGGCTA 980  
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCCTAGA TGCATACAAC 1050  
 CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAACAA GACCAATAAC AATAATGGGA TTAAGCCAT 1120  
 20 AATAAAAAAGT CTCCTAGCAA AGAGAAAGCC AGGACCAAT GGCTTCCTGT CTGGATTTTA CCAATCATTT 1190  
 AAAGAAGAT GAATTCAT CTACTCAAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACTCAT 1260  
 TCTACATGCC CAGTATTACC CTGATTCCAA AACACAGCAA AACACATCA AAAACAAACA AACAAAAAAA 1330  
 CAGAAAGAAA GAAACTACA GGCCAATATC CCTGATGAAT ACTGATACAA AATCCTCAA CAAAACATA 1400  
 GCAAAACCAA TTAACAACA CCTTCGAAAG ATCATTCAAT GTGATCAAGT GGGATTTATT CCAGGGATGG 1470  
 25 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT AAAAAACTA 1540  
 TATGATTATT TCACTTTATG CAGAAAAGC ATTTGATAAA ATTCTGCACC CTTCATGATA AAAACCTCA 1610  
 AAAAACCCAG TATACAGAA ACATACAGCC CAGGCACAGT GGCTCACACC TGCATCCCA GCACTCTGG 1680  
 AGGCCAAGGT GGGATGATT GTTGGGCCCA GGAGTTTGAG ACTAGCCTGG GCAACAAAT GAGACTGTG 1750  
 CTACAAAAA CTTTTTAA AAATTAGCCA GGCATGATGG CATATGCCGT TAGTCCAGC TAGTCTGGAG 1820  
 GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTGAGGC TGCAGTGAGC CATGAACATG TCACTGTACT 1890  
 30 CCAGCCTAGA CAACAGAAACA AGACCCCACT GAATAAGAA AGGAGAGAG AGAAGGGAGA AGGAGGGGAG 1960  
 AAGGGAGGAG GAGGAGAGAG AGGAGGTGGA GGAGAAAGT AAGGGGAAGG GGAAGGGAAA GAGGAGAGAG 2030  
 AAGAAACATA TTTCAACATA ATAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAAGCTGA 2100  
 AGCCTTTCC TCAAGATCTG GAAATGACA AGGGCCCACT TTCACCACTG TATTCAACA TAGTACTAGA 2170  
 AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAAGGA AGAAGTCAA 2240  
 35 TTATCTGT TTGAGATGAT ATGATCTTAT ATCTGGAAGA GACTTAAGAC ACCACTAAAA AACTATTAGA 2310  
 GCTGAAATTT GGTACAGCAG GATACAAAAT CAATGTACAA AAATCAGTAG TATTTCTATA TTCCACAGC 2380  
 AAACAATCTG AAAAAGAAAC CAAAAAGCA GCTACAAATA AAATTAACA GCTAGGAAT AACCAAGAA 2450  
 GTGAAAGATC TCTACAATGA AAATATAAAA ATGTTGATAA AAGAAATTGA AGAGGGCACA AAAAAAGAA 2520  
 40 AGATATTCCTA TGTTCATGA TTGGAAGAA AAATACTGTT AAATGTCCA TACTACCCAA AGCAATTTAC 2590  
 AAATTCATG CAATCCCTAT TAAAACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTTAAGAT 2660  
 TTGTACAGAA CCACAAAGAA CCCAGATAG CCAAGCTAT CCTGACCAA AAGAACAAA CATGAGCAT 2730  
 CACATTACCT GACTTCAAT TATACTACAA AGCTATAGTA ACCCAAACTA CATGGTACTG GCATAAAAC 2800  
 AGATGAGACA TGGACAGAG GAACAGAAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAATCA 2870  
 45 TTTTGACAA AGGTGCCAAG AACATACTT GGGGAAAAGA TAATCTCTTC AATAAATGGT CTGGAGGAA 2940  
 CTGGATATCC ATATGCAAAA TAACATACT AGAATCTGT CTCTACCAT ATACAAAAGC AAATCAAAAT 3010  
 GGTGAAAGG CTTAAATCTA AAACCTCAAA CTTTGCAACT ACTAAAGAA AACACCGGAG AAACCTCTCA 3080  
 GGACATTGGA GTGGGCAAG ACTTCTTGAG TAATCCCTG CAGGCACAGG CAACCAAAGC AAAACAGAC 3150  
 50 AAATGGGATC ATATCAAGTT AAAAGCTTC TGCCAGCAA AGGAAACAT CAACAAAGAG AAGAGACAAC 3220  
 CCACAGAAAT GGAGAAATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATACCAGTA TATATAAGGA 3290  
 GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAT AAGCAAAAGA TCTGGGTAGA 3360  
 CATTTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAATATGTG CTCAACACCA CTGATCATCA 3430  
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTCAT CCCAGTAAA ATGGCTTTTA TTTCAAGAC 3500  
 55 AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGA 3570  
 TTGCTACCA TATGGAGAAC AGTTTGAAAG TTCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC 3640  
 CCATTGCTAG GTATATACCT CAAAAAGGG AATCAGTGA TCAACAAGCT ATCTCCACTC CCACATTAC 3710  
 TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAAAAG 3780  
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA 3850  
 CAGCATGGGG GGCACCTGGT AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT 3920  
 60 CTCCTTACT TGTGGGAGCA AAAATTAAAA CAATTGACAT AGAAATAGAG GAGAAATGGT GTTCTAGAGG 3990  
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTTAA AATAACTAAA CCTGATGTA 4060  
 TTATTACACA TTGATGGCT GTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCTTA CTATATTAAA 4130  
 AATTAATAAT TTAATGGCCA GGCACGGTGG CTATGTCCG TAATCCAGC ACTTTGGGAG GCCGAGGCGG 4200  
 65 GTGGATCACC TGAGGTCAGG AGTTTGAAC CAGTCTGGCC ACCATGATGA AACCTGTCT CTACTAAAGA 4270  
 TACAAAAATT AGCCAGGCGT GGTGGCACAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA 4340  
 TTGCTTGAAC CTGGGAGCGG GAGGTTGCAG TGAGCCGAGA TCATGCCACT GCCTGCGAGC CTGGGTGACA 4410  
 GAGCAAGACT CCATCTCAAA ACAAAACAA AAAAAAGAA ATTAATAATG TAATTTTAT GTACCGTATA 4480  
 70 AATATATACT CTACTATATT AGAAGTTAAA AATTAAAAA ATTATAAAG GTAATTAACC ACTTAATCTA 4550  
 AAATAAGAAC AATGTATGTG GGTGTTCTAG CTCTGAAAG AGTAAAAAGT ATGGCCACGA TGGCAGAAAT 4620  
 GTGAGGAGGG AACAGTGAA GTTACTGTG TTAGCGCTC ATACTCTCTG TAAGTGACTT AATTTTAACC 4690  
 AAAGACAGGC TGGGAGAGT TAAAGAGGCA TTCTATAAGC CCTAAAAACA CTGCTAATAA TGGTGAAGG 4760  
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA 4830  
 CACCGTCTCT TCATTACCAA TGCTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TCGGTTTGTG 4900  
 75 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAAATGGT GGTGATTTC TCCAGAGAA TTAGAGTACC 5040  
 TGGCAGGAAG .CAGGTGGCTG TGTGGACCTG AGCCACTTCA ATCTTCAAG GTCTCTGGCC AAGACCCAGG 5110

	TGCAAGGCAG	AGGCCTGATG	ACCCGAGGAC	AGGAAAGCTC	GGATGGGAAG	GGGCGATGAG	AAGCCTGCCT	5180
	CGTTGGTGAG	CAGCGCATGA	AGTGCCCTTA	TTTACGCTTT	GCAAGATTG	CTCTGGATAC	CATCTGGAAA	5250
	AGGGGGCCAG	CGGGAATGCA	AGGAGTCAGA	AGCCTCCTGC	TCAAACCCAG	GCCAGCAGCT	ATGGCGCCCA	5320
5	CCCGGGCGTG	TGCCAGAGGG	AGAGGAGTCA	AGGCACCTCG	AAGTATGGCT	TAAATCTTTT	TTTACCTGTA	5390
	AGCAGTGACC	AAGGTGTATT	CTGAGGGGAG	CTTGAGTTAG	GTGCTTCTTT	TAAACAGAA	AGTCATGGAA	5460
	GCACCTTCT	CAAGGGAATA	CCAGAGCCCC	GCTCTGCGGT	CATTACCTC	TTTCTCTCT	CCCTCTCTTG	5530
	CCCTCGCGGT	TTCTGATCGG	GACAGAGTGA	CCCCGTGGA	GCTTCTCGA	GCCCGTGCTG	AGGACCCCTCT	5600
	TGCAAGAGGC	TCCACAGACC	CCCGCCCTGG	AGAGAGGAGT	CTGAGCCTGG	CTTAATAACA	AACTGGGATG	5670
	TGGCTGGGGG	CGGACAGCGA	CGGCGGGATT	CAAGACTTA	ATTCCATGAG	TAAATTCAC	CTTCCACAT	5740
10	CCGAATGGAT	TTGGATTTTA	TCTTAATATT	TTCTTAAATT	TCATCAAATA	ACATTACAGG	CTGCAGAAAT	5810
	CCAAAGGCGT	AAAACAGGAA	CTGAGCTATG	TTTGCCAAAG	TCCAAGGACT	TAATAACCAT	GTTCAGAGGG	5880
	ATTTTTCCGC	CTAAGTACTT	TTTATTGGTT	TTTATAAGGT	GGCTTAGGGT	GCAAGGGAAA	GTACACGAGG	5950
	AGAGGCCTGG	GCGGCAGGGC	TATGAGCAGC	GCAGGGCCAC	CGGGGAGAGA	GTCCCGGGCC	TGGGAGGCTG	6020
	ACAGCAGGAC	CACCTGACCGT	CCTCCCTGGG	AGCTGCCACA	TTGGGCAACG	GGAAGGCGGC	CACGCTGCGT	6090
15	GTGACTCAGG	ACCCCATACC	GGCTTCTGGG	CCCCCACCAC	ACTAACCCAG	GAAGTCACGG	AGCTCTGAAC	6160
	CCGTGGAAAC	GAACATGACC	CTTGCTCGCC	TGCTTCCCTG	GGTGGGTCAA	GGGTAAATGAA	GTGGTGTGCA	6230
	GGAAATGGCC	ATGTAAATTA	CACGACTCTG	CTGATGGGGA	CCGTTCTCTC	CATCATTATT	CATCTTACCC	6300
	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TCGGGTGTGA	CAAGCCATGA	CAAAACTCAG	TACAACACCC	6370
20	ACTCTTTTAC	TAGGCCACCA	GAGCAGGSC	CACACCCTTG	ATATATTAAG	AGTCCAGGAG	AGATGAGGCT	6440
	GCTTTCAGCC	ACCAGGCTGG	GGTGACAACA	CGGGCTGAAC	AGTCTGTTC	TCTAGACTAG	TAGACCCCTG	6510
	CAGGCAGTCC	CCCAGATTCT	AGGGCTCTGT	TGCTGTCTCC	CGAGGGCGCC	ATCTGCCCTG	GAGACTCAGC	6580
	CTGGGGTGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCTT	CCGCTCCAG	GCCTCAGCTT	TCTCAGCAGC	6650
	TTCTTAAACC	CTGGGTGGGC	CGTGTTCAG	CGCTACTGTC	TCACCTGTCC	CACCTGTGCT	TGTCTCAGCG	6720
25	AGTAGCTCG	CACGGTTCCT	CCTCAGATGG	GGTGTCTGTC	TCCTTCCCCA	ACACTCAGAT	GGCTTGAAGG	6790
	GAGGAGATTC	TGCGCCTCCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCCC	GATGAGGTTT	6860
	CCTGGCGTCC	GGCTGCACGC	TGACCTCCAT	TTCCAGGCGC	TCCCGCTCTC	CTGTCTCTG	CCGGGCGCTG	6930
	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCCTT	TCCACGTCCA	GCTGCGGTG	TCTCTGCCCC	CTAGGGTCTC	7000
	GGGGTTTTTA	TAGGCATAGG	ACGGGGCGGT	GGTGGGCCAG	GGCGCTCTTG	GGAAATGCAG	CATTTGGGTG	7070
30	TGAAAGTAGG	AGTGCCCTGTC	CTCACCTAGG	TCCACGGGCA	CAGGCCTGGG	GATGGAGGCC	CCGCGAGGGA	7140
	CCCGGCCCTC	TCTGCCACAG	ACTTCTCTGC	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	TCCACAGC	7210
	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATTT	GCCCCACAGC	CTGGGAATT	7280
	CACGTGACTA	CGCACATCAT	GTACACACTC	CCGTCCACGA	CCGACCCCGC	CTGTTTATT	TAAATAGCTA	7350
	CAAAAGCAGG	AAATCCCTGC	TAAATGTCC	TTTAAACAA	TGGTTAAACA	AACGGGTCCA	TCCGACCGGT	7420
35	GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATA	AGCCTGCAGG	CATCTCAAGG	GAATTACGCT	7490
	GAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TACGCAACAT	GCTCAAAAG	AAAGAAATTC	ACCCATGGC	7560
	AGGGGAGTGG	TTAGGGGGTG	TAGGACCGGT	GGGGGCGGCA	GCTGGGGGCT	ACTGCACGCA	CCTTTTACTA	7630
	AAGCCAGTTT	CCTGGTCTG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	TGGGATGGG	7700
	GGAAACCCGA	GGCTGTGCCA	TCCTTGCCAT	GGCCGAGTGT	CCTGGGCAGG	ATAATGCTCT	AGAGATGCCC	7770
40	AGCTCTGAT	TCCCCCAAA	CTGTGGACAG	AACCCGCCCC	GCCCCAGGGC	CTTTCAGGTT	GTGATCTCCG	7840
	TGAGGACCTT	GAGGTCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCCGAAAA	GTAATCCAGG	GGTTCGGGGA	7910
	AGAGGCGGGC	AGGAGGGTCA	GAGGGGGGCA	GCCTCAGGAC	GATGGAGGCA	GTCAGTCTGA	GGCTGAAAAG	7980
	GAGGGGAGGG	CCTCGAGCCC	AGGCCTGCAA	GCGCCTCCAG	AAGCTGGAAA	AAGCGGGGAA	GGGACCCCTC	8050
	ACGGAGCCCTG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	TCCGGCTCC	8120
45	GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTGCAGAA	GCAACAGGAA	8190
	ACCCATGCAC	TGTGAATCTA	GGATTATTTT	AAAACAAAGG	TTTACAGAAA	CATCCAAGGA	CAGGGCTGAA	8260
	GTGCCCTCCG	GCAAGGGCAG	GGCAGGCACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	TACTTACTTT	8330
	CTGAGACAGA	GTTATGCTCT	TGTTGCCACG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	CTGCAACCTC	8400
	CGTCTCTTGG	GTTCAGCAAA	TCTCGTGCC	TCAGCCTCCC	AAGTAGCTGG	GATTTTCAGG	GCTGCAACCC	8470
50	ACACCCGGCT	AATTTGTAT	TTTTAGTAGA	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	TCTCAAAATC	8540
	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	CACCTGCACCT	8610
	GGCCTATTTA	ACCATTTTAA	AACTTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	CATGGAGTTC	8680
	AATTTCCCTT	TACTCAGGA	GTTACCCCTC	TTTGATATTT	TCTGTAATTC	TTCGTAGACT	GGGGATACAC	8750
	CGTCTCTTGA	CATATTCA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCACCTGCA	GGGCAGCTGG	8820
55	GAGGCTGCAG	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	ATCAGGGGCG	8890
	AAGTGTGGAC	ACTGTCTCTG	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	GTAGAAATTA	AAGTCCATCC	8960
	CTCCTACTCT	ACTGGGATTG	AGCCCTTTC	CTATCCCCC	CCAGGGGCG	AGGAGTTTCT	CTCACTCCTG	9030
	TGGAGGAAGG	AATGATACTT	TGTTATTTTT	CACCTGCTGT	ACTGAATCCA	CTGTTTCATT	TGTTGGTTTG	9100
	TTTGTTTTGT	TTTGAGAGGC	GGTTTCACTC	TTGTTGCTCA	GGCTGGAGGG	AGTGCAATGG	CGCGATCTTG	9170
60	GCTTACTGCA	GCCTCTGCCT	CCCAGGTTCA	AGTGATTCTC	CTGCTTCCGC	CTCCCATTTG	GCTGGGATTA	9240
	CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGTATTTT	TAGTAGAGAC	GSGGGTGGGT	GGGGTTCAAC	9310
	ATGTTGGCCA	GGCTGCTCTC	GAACCTTCTG	CCTCAGATGA	TCCACCTGCC	TCTGCTCTCT	AAAGTGTCTG	9380
	GATTACAGGT	GTGAGCCACC	ATGCCAGCT	CAGAATTTAC	TCTGTTTAGA	AACATCTGGG	TCTGAGGTAG	9450
65	GAAGCTCACC	CCACTCAAGT	GTGTGGTGT	TTTAAGCCAA	TGATAGAAAT	TTTTTATTGT	TGTTAGAAC	9520
	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	TGCACCCATA	9590
	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGAGG	CGTTTCTCTG	CCATGCCATC	9660
	GGTGTAAAT	ACTCCAGCAT	AATCTTCTGC	TTCCATTCTT	TCTCTTCCCT	CTTTTAAAT	TGTTTCTTCT	9730
	ATGTTGGCTT	CTCTGCAGAG	AACCACTGTA	AGCTACAAC	TAACTTTTGT	TGGAACAAAT	TTTCCAAACC	9800
	GCCCTTTTGC	CCTAGTGCCA	GAGACAATTC	ACAAACACAG	CCCTTTAAAA	AGGCTTAGGG	ATCACTAAGG	9870
70	GGATTCTAG	AAGAGCGACC	TGTAATCCTA	AGTATTTACA	AGACGAGGCT	AACCTCCAGC	GAGCGTGACA	9940
	GCCACGGGAG	GGTGCGAGGC	CTGTTCAAAT	GCTAGCTCCA	TAAATAAAGC	AATTTCTCTC	GGCAGTTTCT	10010
	GAAAGTAGGA	AAGGTTACAT	TTAAGGTTGC	GTTTGTAGC	ATTTCAGTGT	TTGCCGACCT	CAGCTACAGC	10080
	ATCCCTGCAA	GGCCTCGGGA	GACCCAGAAG	TTTCTCGCCC	CCTTAGATCC	AAACTTGAGC	AACCCGAGGT	10150
	CTGGATTCTT	GGGAAGTCTT	CAGCTGTCTT	GCGGTTGTGC	CGGGGCCCCA	GGTCTGGAGG	GGACGAGTGG	10220
75	CCGTGTGGCT	TCTACTGCTG	GGCTGGAAGT	CGGSCCTCCT	AGCTCTGCG	TCCGAGGCTT	GGAGCCAGGT	10290
	GGCTGGACCC	CGAGGCTGCC	CTCCACCTGT	TGCGGGCGGG	ATGTGACCAG	ATGTTGGCCT	CATCTGCCAG	10360
	ACAGAGTGCC	GGGGCCACAG	GTCAAGGCCG	TTGTGGCTGG	TGTGAGGCGC	CCGGTGGCGG	GCCAGCAGGA	10430
	GCGCCTGGCT	CCATTTCCCA	CCCTTTCTCG	ACGGGACCGC	CCCGTGGGTT	GATTAACAGA	TTTGGGGTGG	10500

	TTTGCTCATG	GTGGGGACCC	CTCGCCGCT	GAGAACCTGC	AAAGAGAAAT	GACGGGCTTG	TGTCAAGGAG	10570
	CCCAAGTCGC	GGGGAAGTGT	TGCAGGGAGG	CACTCCGGGA	GGTCCCGCGT	GCCCGTCCAG	GGAGCAATGC	10640
	GTCTCTGGGT	TCGTCCCCAG	CCGCGTCTAC	GCGCCTCCGT	CCTCCCCCTC	ACGTCCGGCA	TGTCTGGTGC	10710
	CCGGAGCCCG	ACGCCCCGCG	TCCGGACCTG	GAGGCAGCCC	TGGGTCTCCG	GATCAGGCCA	GCGGCCAAAG	10780
5	GGTCGCCGCA	CGCACCTGTT	CCCAGGGCCT	CCACATCATG	GCCCCCTCCT	CGGGTTACCC	CACAGCCTAG	10850
	GCCGATTGCA	CCTCTCTCCG	CTGGGGCCCT	CGCTGGCGTC	CCTGCACCTT	GGGAGCGCGA	GCGCGCGCGC	10920
	GCGGGGGAAG	CGCGGCCAG	ACCCCCGGT	CGCCCCGAG	CAGCTGCGCT	GTCGGGGCCA	GCGCGGCTC	10990
	CCAGTGGATT	CGCGGGCACA	GACGCCCAGG	ACCGCGCTCC	CCACGTGGCG	GAGGAGTGGC	GGACCCGGGC	11060
	ACCCGTCTCT	CCCTTTCACC	TTCAGCTCC	GCCTCCTCCG	CGCGGACCCC	GCCCCGTCCC	GACCCCTCCC	11130
10	GGGTCCCCGG	CCCAGCCCCC	TCCGGGCCCT	CCCAGCCCCC	CCCTTTCCTT	TCCGCGGCCC	CGCCCTCTCC	11200
	TGCGGGCGCG	AGTTTCAGGC	AGCGCTGCGT	CCTGCTGCGC	ACGTGGGAAG	CCCTGGCCCC	GGCCACCCCC	11270
	CGGATGCGCG	GCCTCCCGG	CTGCCGAGCC	GTGCGCTCCC	TGCTGCGCAG	CCACTACCGC	GAGGTGCTGC	11340
	CGCTGGCCAC	GTTCTGTGCG	CGCCTGGGGC	CCAGGGCTGT	GCGGTGTGTG	CAGCGCGGGG	ACCCGGCGGC	11410
	TTTCCGCGCG	CTGGTGGCCC	AGTGCTTGGT	GTGCGTGCCC	TGGGACGCAC	GCGCGCCCCC	CGCGCCCCC	11480
15	TCCTTCCGCC	AGGTGGGCTT	CCCCGGGCTC	GGCGTCCGGC	TGGGTTGAG	GCGCGCCGGG	GGGAACCAGC	11550
	GACATCGCGA	GAGCAGCGCA	GGCGACTCAG	GGCGCTTCCC	CCGCAGGTGT	CCTGCCTGAA	GGAGCTGGTG	11620
	GCCCGAGTGC	TGCAGAGGCT	GTGCGAGCGC	GGCGCGAAGA	ACGTGCTGGC	CTTCGGCTTC	CGCTGCTGGG	11690
	ACGGGGCCCG	CGGGGGCCCC	CCCCAGGCCT	TCACACCAGC	CGTGGCGAGC	TACCTGCCCA	ACACGGTGAC	11760
	CGACGCACATG	CGGGGAGGCG	GGGCGTGGGG	GCTGCTGCTG	GCGCGCGTGG	GCGAGCAGCT	GCTGGTTTCA	11830
20	CTGCTGGCAC	GCTGCGCGCT	CTTTGTGCTG	GTGCTTCCCA	GCTGCGCCTA	CCAGGTGTGC	GGGCGCGCGC	11900
	TGTACCAAGCT	CGCGCTTGCC	ACTCAGGCCC	GGCCCCCGCC	ACACGCTAGT	GGACCCCGAA	GGCGCTTGGG	11970
	ATCGGAACGG	GCTTGGAAAC	ATAGCGTCAG	GGAGGCCGGG	GTCCCCCTGG	GCCTGCCAGC	CCCGGTGCGG	12040
	AGGAGGCGCG	GGGGCAGTGC	CAGCCGAAGT	CTGCCGTGTC	CCAAGAGGCC	CAGGCGTGGC	GCTGCCCTGT	12110
	AGCCGGAGCG	GACGCCCGTT	GGGCAGGGGT	CCTGGGCCCA	CCCGGGCAGG	ACGCGTGGAC	CGAGTGACCG	12180
25	TGCTTTCTGT	CTGGTGTAC	CTGCCAGACC	CGCCGAAGAA	GCCACCTCTT	TGGAGGGTGC	GCTCTCTGGC	12250
	ACGCGCCACT	CCCACCCATC	CGTGGGCCGC	CAGCACACAG	CAGGCCCCCC	ATCCACATCG	CGGCCACAC	12320
	GTCCCTGGGA	CACGCTTGT	CCCCGGTGT	ACGCCGAGAC	CAAGCACTTC	CTCTACTCCT	CAGCGACAA	12390
	GGAGCAGCTG	CGGCCCTCCT	TCCTACTCAG	CTCTCTGAGG	CCCAGCCTGA	CTGGCGCTCG	GAGGCTCGTG	12460
30	GAGACCATCT	TTCTGGGTTT	CAGGCCCTGG	ATGCCAGGGA	CTCCCCGAG	GTTGCCCCGC	CTGCCCTCAG	12530
	GCTACTGGCA	AATGCGGCC	CTGTTTCTGG	AGCTGCTTGG	GAACCAACGC	CAGTGGCCCT	ACGGGGTGT	12600
	CCTCAAGACG	CAGTGGCCCG	TGCGAGCTGC	GCTCACCCCA	GCAGCCGGTG	TCTGTGCCCG	GGAGAAGCCC	12670
	CAGGGCTCTG	TGGCGGCCCC	CGAGGAGGAG	GACACAGACC	CCCGTCCGCT	GCTGACAGTG	CTCCGCCAGC	12740
	ACAGCAGCCC	CTGGCAGGTG	TACGGCTTCG	TGCGGGCCCT	CCTGCGCCCG	CTGGTGGCCC	CAGGCCCTCT	12810
35	GGGCTCCAGG	CACAACGAAC	GCCGCTTCTT	CAGGAACACC	AAGAAGTTCA	TCTCCCTGGG	GAAGCATGCC	12880
	AAAGCTCTGC	TGCAGGAGCT	GACGTGGGAG	ATGAGCGTGC	GGGACTGCGC	TGGGTGCGC	AGGAGCCAGC	12950
	GTAGGAGGT	GCTGGCCGTC	GAGGGCCGAG	GCCCCAGAGC	TGAATGCAGT	AGGGGCTCAG	AAAGGGGGGC	13020
	AGGCAGAGCC	CTGCTCTCC	TGCTTCCATC	GTCACGTGGG	CACACGTGGC	TTTTCGCTCA	GGACGTGCGG	13090
	TGGACACGGT	GATCTCTGCC	TCTGCTCTCC	CTCCTGTCCA	GTTTGATATA	ACTTACGAGG	TTCACCTTCA	13160
40	CGTTTGTATG	GACACGCGGT	TTCCAGGCGC	CGAGGCCAGA	GCAGTGAACA	GAGGAGGCTG	GGCGCGCGAG	13230
	TGGAGCCGGG	TGCGCCGCAA	TGGGGAGAAG	TGCTTGGGAG	CACAGACGCT	CTGGCGAGGG	TGCTGCGAGG	13300
	TTACCTATAA	TCCTCTTCGC	AATTTCAAGG	GTGGGAATGA	GAGGTGGGGA	CGAGAACCCT	CTCTTCTGCG	13370
	GGGTGGGAGG	TAAGGGTTTT	GCAGGTGCAC	GTGGTCAGCC	AATATGCAGG	TTTGTGTTTA	AGATTTAATT	13440
	GTGTGTTGAC	GGCCAGGTGC	GCTGGCTCAC	GCCGGTAATC	CCAGCACTTT	GGGAAGCTGA	GGCAGGTGGA	13510
45	TCACCTGAGG	TCAGGAGTTT	GAGACCAACC	TGACCAACAT	GGTGAAACCC	TATCTGTACT	AAAAATACAA	13580
	AAATTAGCTG	GGCATGTTGG	TGTGTGCTGT	TAATCCAGCC	TACTTGGGAG	GCTGAGGCGC	GAGAATCACT	13650
	TGAACCCAGG	AGGCGGAGGC	TGCAGTGAGC	TGAGATTGTG	CCATTGTACT	CCAGCCTGGG	GCACAAGAGT	13720
	GAAACTCTGT	CTTTAAAAAA	AAAAAGTGT	CGTTGATTGT	GCCAGGACAG	GATAGAGGGA	GGGAGATAAG	13790
	ACTGTCTCTC	AGCACAGATC	CTGGTCCCAT	CTTTAGGTAT	GAAGAGGGCC	ACATGGGAGC	AGAGGACAGC	13860
50	AGATTGGCTC	ACCTGCTGAG	GAAGGGACAG	TGTTGTGGGG	TGTTAGGGGG	ATGGTGTGTC	TGGGCCCTGC	13930
	CGTGTCCCCA	CCCTGTTTTT	CTGGATTGGA	TGTTAGGAA	CCTCCGCTCC	AGCCCCCTTT	TGGCTCCGAG	14000
	TGCTCCACAG	CCCTACCGTG	GCAGCTAGAA	GAAGTCCCGA	TTTCACCCCT	TCCCCACAAA	CTCCCAAGAC	14070
	ATGTAAGACT	TCCGGCCATG	CAGACAAAGG	GGGTGACCTT	CTTGGGGCTC	TTTTTTTCTT	TTTTTCTTCT	14140
	TTATGGTGGC	AAAAGTCATA	TAACATGAGA	TTGGCACTCC	TAACACCGTT	TTCTGTGTAC	AGTGCAGAAT	14210
55	TGCTAACTCG	GCGGTGTTTA	CAGCAGGTTC	CTTGAAATGC	TGCGTCTTGC	GCTGACTGGA	GTCCCTACCC	14280
	ATCGAACGGC	AGCTGCCTCA	CACCTGCTGC	GGCTCAGGTG	GACCAACGCG	AGTCAGATAA	GGGTCAATGA	14350
	ACCCAGTTTT	GCTTTTTGTG	CTCCAGCTTC	CTTCGTTGAG	GAGAGTTTGA	GTTCTCTGAT	CAGGACTCTG	14420
	CCTGTCAATT	CTGTCTCTGT	ACTTCAGATG	AGGTCACAAT	CTGCCCCCTG	CTTATGCAGG	GAGTGAGGCG	14490
	TGGTCCCCCG	GTGTCCCTGT	CACGTGCAGG	GTGAGTGAGG	CGTTGCCCCC	AGGTGTCCCT	GTCAGGTGTA	14560
60	GGGTGAGTGA	GGCGCGGCC	CCGGGTGTCC	CTGTCCCGTG	CAGCGTGATT	GAGGTGTGGC	CCCCGGGTGT	14630
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCGCC	ATCCCCGGGT	GTCCTGTGTA	CGTGTAGGGT	GAGTGAGGCG	14700
	TGGTCCCCCG	GTGTCCCTGT	CCCGTGCAGG	GTGAGTGAGG	CACGTGCCCC	GGGTGTCCCT	GTCACGTGCA	14770
	GGGTGAGTGA	GGCGCGGTCC	CCGGGTGTCC	CTCTCAGGTG	TAGGGTGAGT	GAGGCGCGGC	CCGAGGGTGT	14840
	CGCTGTACAG	TGTAGGGTGA	GTGAGGCACC	GTCCTGGGT	GTCCTCCCA	GGTATAGGGT	GAGTGAGGCA	14910
65	CTGTCCCCCG	GTGTCCCTGT	CACGTGCAGG	GTGAGTGAGG	CGCGGCCCCC	GGGTGTCCCT	CTCAGGTGCA	14980
	GGGTGAGTGA	GGCGCTGTCC	CTGGGTGTCC	CTGTCTCTGT	TAGGGTGAGT	GAGGCTCTGT	CCCCAGGTGT	15050
	CCTTGGCGTT	TGCTCACTTG	AGCTTGTCTC	TGAATGTTTG	CTCTTTCTAT	AGCCACAGCT	GCGCGGGTTG	15120
	CCCATTTGCT	GGGTAGATGG	TGCAGGCGCA	GTGCTGGTCC	CCAAGCCTAT	CTTTTCTGAT	GCTCGGCTCT	15190
	TCTTGGTCAC	CTCTCCGTTT	CATTTTGCTA	CGGGGACACG	GGAGTGCAGG	CTCTCGCCTC	CCCGTGGCCA	15260
70	GGCACTGCAG	CCACAGCTTC	AGGTCCGCTT	GCCTCTGTTG	GGCCTGGCTT	GCTCACACAG	TGCCCCGCAC	15330
	ATGCATGCTG	CCAATACTCC	TCTCCAGCT	TGCTCTATGC	CGAGGCTGGA	CTCTGGGCTG	CCTGTGCTGT	15400
	CTGCCACGTG	TTGCTGGAGA	CATCCAGAA	AGGGTTCTCT	GTGCCCTGAA	GGAAAGCAAG	TCACCCAGGC	15470
	CCCTCACTTT	GTCTGTTTTT	CTCCCAAGCT	CGCCCTCTGC	TTGGGCCCTT	TGGGTGGGTT	GCAACGCTTG	15540
	TCACCTTATT	CTGGGCACCT	GCCGCTCAT	GCTTAGGCTG	GGCTCTGCTT	CCAGTGGCCC	CCTCACATGG	15610
75	ATTGACGTCC	AGCCACAGGT	TGGAGTGTCT	CTGTCTGTCT	CTGTCTGTA	GACCCACGTG	GAGGGCCGGT	15680
	TTCTCGGCA	GCCTTCGTCA	GACTTCCCTC	TTGGGTCTTA	GTTTGAATTT	TACTGATTT	ACCTCTGACG	15750
	TTTCTATCTC	TCCATTGTAT	GCTTTTTCTT	GGTTTATTCT	TTCTTCTCTT	TTCTAGCTTC	TGAGTTTATG	15820
	CATGCCCTTC	CCTCTAAGTG	CTGCCTTACC	TGCACCTCTG	GTTTGTGATG	GAAGTAATCT	CAACATCAGC	15890

	CACTTTCAAG	TGTTCTTAAA	ATACTTCAAA	GTGTTAATAC	TTCTTTTAAAG	TATTCTTATT	CTGTGATTTT	15960
	TTTCTTTTGTG	CACGCTGTGT	TTTGACGTGA	AATCATTTTG	ATATCAGTGA	CTTTTAAAGTA	TTCTTTAGCT	16030
	TATTCCTGTGA	TTTCTTTGAG	CAGTGAGTTA	TTTGAACACT	GTTTATGTTT	AAGATATGTA	GAGTATCAAG	16100
5	ATACGATAGAG	TATTTTAAAGT	TATCATTTTA	TTATTGATTT	CTAACTCAGT	TGTGTAGTGG	TCTGTATAAT	16170
	ACCAATTATT	TGAAGTTTGC	GGAGCCTTGC	TTTGTGATCT	AGTGTGTGCA	TGGTTTCCAG	AACGTGCCAT	16240
	TGTAATTTTG	ACATCCTGTG	AATAGTGGGC	ATGCATGTTC	ACTATATCCA	GCTTATTAAAG	GTCCAGTGCA	16310
	AAGCTTCTGT	CTCCTTCTAG	ATGCATGAAA	TTCCAAGAAG	GAGGCCATAG	TCCCTCACCT	GGGGGATGGG	16380
	TCTGTTTCATT	TCTTCTCGTT	TGGTAGCATT	TATGTGAGGC	ATTGTTAGGT	GCATGCACGT	GGTAGAATTT	16450
	TTATCTTCTCT	GATGAGTGAA	TCTTTTGGAG	ACTTCTATGT	CTCTAGTAAT	CTAGTAATTC	TTTTTTTAAA	16520
10	TTGCTCTTAG	TACTGCCACA	CTGGGCTTCT	TTTGATTAGT	ATTTTCTGTC	TGTGCTGTTT	TTCTGGCTTT	16590
	AATTTATATA	TATATATATA	TTTTTTTTTT	TTTTGAGACA	GAGTCTTGGT	CTGTGCGCCA	GGGTGAGTGC	16660
	AGTGGTGTGA	TCACAGGTCA	GTGTAACTTT	TACCTTCTGG	CCTGAGCCGT	CCTCTCACCT	CAGCCTCCTG	16730
	AGTAGCTGGA	ACTGCAGACA	CGCACCCGTA	CACCTGGCTA	ATTTTAAAT	TTTTTCTGGA	GACAGGGTCT	16800
	TGCTGTGTG	CCCAGGCTGG	TCTCAAACTC	TGGACTCAA	GGGATCCATC	TACCTCGGCT	TCCCAAAGTG	16870
15	CTGAATTACA	GGCATGAGCC	ACCATGTCTG	GCCTAATTTT	CAACACTTTT	ATATCTTAT	AGTGTGGGTA	16940
	TGTCCTGTGA	ACAGCATGTA	GGTGAATTTT	CAATCCAGTC	TGACAGTCGT	TGTTTAACTG	GATAACTGTA	17010
	TTTATTTTCA	TTTTTTTGTG	ACTAGAGACC	CGCCTGGTGC	ACTCTGATTC	TCCACTTGCC	TGTTGCATGT	17080
	CCTCGTTCCC	TGTTTCTCA	CCACCTCTTG	GGTGGCATG	TGCGTTTCTT	GCCGAGTGTG	TGTTGATCCT	17150
	CTCGTTGCTC	CCTGGTCACT	GGGCATTGTC	TTTTATTCTT	CTTTGCTTAG	TGTTACCCCT	TGATCTTTTT	17220
20	ATTGTCGTTG	TTTGCTTTTG	TTTATTGAGA	CAGTCTCACT	CTGTCAACCA	GGCTGGAGTG	TAATGTCACA	17290
	ATCTCGGCTC	ACTGCAACCT	CTGGCTCCTC	GGTTCAGACA	GTTCTCATTC	CTCAACCTCA	TGAGTAGCTG	17360
	GGATTACAGG	CGCCCAACCA	CACGCTGGCC	TAATTTTGTG	ATTTTATAGTA	GAGATAGGCT	TTCAACATGT	17430
	TGGCCAGGCT	GGTCTCAAACT	TCCCTGACCTC	AAGTGATCTG	CCCGCCTTGG	CCTCCACACAG	TGCTGGGAT	17500
	ACAGGTGCAA	GCCACCGTGC	CCGGCATACC	TTGATCTTTT	AAAATGAAGT	CTGAAACATT	GTACCCCTTG	17570
25	TCCTGAGCAA	TAAGACCCCT	AGTGATTTT	AGCTCTGGCC	ACCCCCACAG	CTGTGTGCTG	TTTTCCCTGC	17640
	TGACTTAGTT	CTATCTCAGG	CATCTTGACA	CCCCCACAAAG	CTAAGCATT	TTAATATTGT	TTTCCGTGTT	17710
	GAGTGTTTCT	GTAGCTTTTG	CCCCGCCCTG	CTTTTCTCTC	TTTGTCTCCC	GTCTGTCTTC	TGCTCAGGCG	17780
	CCGCGCTCTG	GGGTCCCTTT	CTTGTCTCTT	TGCGTGGTTC	TTCTGTCTTG	TTATTGCTGG	TAAACCCAG	17850
	CTTTACCTGT	GCTGGCCTCT	ATGGCATCTA	GCGACGTCGG	GGGACCTCTG	CTTATGATGC	ACAGATGAAG	17920
30	ATGTGGAGAC	TCACGAGGAG	GGCGGTATC	TTGGCCCGTG	AGTGTCTGGA	GCACACGCTG	GCCAGCGTTC	17990
	CTTAGCCAGT	GAGTGACAGC	AACGTCCGCT	CGGCTGGGTT	TCAGCTCGGA	AAACCCACAG	CATGTCCGGG	18060
	TCTGGTGGCT	CCGCGGTGTC	GAGTTTGAAA	TCGCGCAAAAC	CTGCGGTGTG	CGGCCAGCTC	TGACGGTGCT	18130
	GCCTGGCGGG	GGAGTGCTG	CTTCTCCCTT	CTGCTTGGG	AACCAAGGACA	AAGGATGAGG	CTCCGAGGCC	18200
	TTGTGCGCCA	ACAGGAGCAT	GACGTGAGCC	ATGTGGATAA	TTTTAAATTT	TCTAGGCTGG	GCGCGGTGGC	18270
35	TCAGCCCTGT	AATCCAGCA	CTTTGGGAGG	CCAAGGCGGG	TGGATCACGA	GGTCAGGAGG	TCGAGACCAT	18340
	CCTGGCCAAC	ATGATGAAAC	CCCATCTGTA	CTAATAAACAC	AAAAATTAGC	TGGCGCTGGT	GGCGGTGGCC	18410
	TGTAATCCCA	GCTACTCGGG	AGGCTGAGGC	AGGAGAATTG	CTTGAACCTG	GGAGTTGGAA	GTTCGAGTGA	18480
	GCCGACATTG	CACCACTGCA	CTCCAGCCTG	GCAACACAGC	GAGACTCTGT	CTCAAAAAAA	AAAAAAAAAA	18550
40	AAAAAATAAA	AATTCTAGTA	GCCACATTAA	AAAAAGTAAA	AAGAAAAGGT	GAAATTAATG	TAATAATAGA	18620
	TTTTACTGAA	GCCACGATG	TCCACACCTC	ATCATTTTAG	GGTGTATTG	GTGGGAGCAT	CACCTACAGG	18690
	ACATTTGACA	TTTTTTGAGC	TTTGTCTCGG	GGATCCCTGT	TGTAGGTCCC	GTGCGTGGCC	ATCTCGGGCT	18760
	GGACCTGCTG	GGCTTCCCAT	GGCATAGGCT	GTTGTACACG	ATGGTGACAG	TCCGGGATGA	GGTCCGACAG	18830
	CCCTCAGTGA	GCTGGATGTG	CAGTGTCCGG	ATGGTGACAG	TCTGGGATGA	GGTCCGACAG	CCCTGCTGTG	18900
45	AGCTGGATGT	GTGGTGCTG	GATGGTGAGC	GTGAGGGGTG	AGGTCTCCAG	GCCCTCGGTC	AGCTGAGGTC	18970
	ATGGAGTCCG	GATGATGAGC	GTCCGGGGTG	AGGTGCGCCAG	GCCCTGCTGT	GAGCTGGATG	TGTGGTGTCT	19040
	GGATGGTGCA	GGTCAGGGGT	GAGGTCTCCA	GGCCCTCGGT	AAGCTGGAGG	TATGGAGTCC	GGATGATGCA	19110
	GGTCCGGGTT	GAGGTGCGCA	GGCCCTGCTG	TGAGCTGGAT	GTGTGGTGTG	TGGATGGTGC	AGGTCTGGGG	19180
	TGAGGTGACC	AGGCCCTGCG	GTGAGCTGGG	TGTGCGGGTG	CTGGATGGTG	CAGGTCTGGA	GTGAGGTCTG	19250
	CAGACCGTGC	CAGACCATGC	GGTGAGCTGG	ATATGCGGTG	TCCGGATGGT	GCAGGTCTGG	GGTGAGGTTC	19320
50	CCAGGCCCTG	CTGTGAGTTG	GATGTGGGTT	GTCCGGATGC	TGCAGGTCCG	GTGTGAGGTC	ACCAGGCCCT	19390
	GCTGTGAGCT	GGATGTGTGG	TGTCTGGATG	GTGACAGTCT	GGGTGGAAGG	TCCGCCAGGC	CTCGCTTGTG	19460
	AGCTGGATGT	GTGGTGTCTG	GATGGTGACG	GTCTGGAGTG	AGGTGCGCCAG	GCCCTCGGTC	AGCTGGATGT	19530
	GCAGTGTCCA	GATGGTGACG	GTCCGGGGTG	AGGTGCGCCAG	ACCCTGCGGT	GAGCTGGATG	TCCGGTGTCT	19600
	GGATGGTGCA	GGTCTGGAGT	GAGGTGCGCA	GGCCCTCGGT	GAGCTGGATG	TATGGAGTCC	GGATGGTGCC	19670
55	GGTCCGGGGT	GAGGTGCGCA	GACCTGCTGT	TGAGCTGGAT	GTGCGGTGTC	TGGATGGTAC	AGGTCTGGAG	19740
	TGAGGTGCGC	AGACCTGTCT	GTGAGCTGGA	TATGCGGTGT	CCGGATGGTG	CAGGTGAGGG	GTGAGGTCTC	19810
	CAGGCCCTCG	GTGAGCTGGA	GGTATGGAGT	CCGGATGATG	CAGGTCCGGG	GTGAGGTGCG	CAGGCCCTCG	19880
	TGTGAACCTG	ATGTGCGGGC	TCTGGATGGT	GCAGGTCTGG	GGTGTGGTGC	CCAGGCCCTC	GGTGAGCTGG	19950
	AGGTATGGAG	TCCGGATGAT	GCAGGTCCGG	GGTGAGGTTC	CCAGGCCCTG	CTGTGAGCTG	GATGTGCGGC	20020
60	GTCTGGATGG	TGCAGGTCTG	GGGTGTGGTG	GCCAGGCCCT	CGGTGAGCTG	GAGGTATGGA	GTCCGGATGA	20090
	TGCAGGTCCG	GGGTGAGGTT	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGAGTCCG	20160
	GGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGAGGTC	GGGTGAGGTC	20230
	CACCAGGCCCT	TCCGGTGAGC	TGGTTGTGCG	GTGTCCGGTT	GCTGCAGGTC	CGGGGTGAGT	TCCGCCAGGC	20300
	CTCGGTGAGC	TGGATGTGCG	GTGTCCCGCT	GTCCGGATGG	TGCAGGTCCA	GGGTGAGGTC	GCTAGGCCCT	20370
65	TGGTGGGCTG	GATGTGCGGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGAGCTG	20440
	GATGTGCGGT	GTCTGCATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATGTGTGGT	20510
	GTCCGGATGG	TGCAGGTCCG	CGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCGG	TGCTGGATG	20580
	GGGTGAGGTC	GGGTGAGGTC	AGCCCAAGGC	TTCGGTGAGC	TGGATGTGCG	GTGTCCGGAT	GGTGACAGTC	20650
	CGGGGTGAGG	TGCCAGGCC	CTCGGTTAG	CTGGATATGC	GGTGTCCGGA	TGGTGACAGT	CGCGGGTGA	20720
70	GTCAACAGGC	CCTGCGGTTA	GCTGGATGTC	CGGTGTCTGG	ATGGTGACAG	TCCGGGGTGA	GGTCCGCCAG	20790
	CCCTGCTGTG	AGCTGGATGT	GCTGTATCCG	GATGGTGACG	GTCCGGGGTG	AGGTGCGCAG	GCCCTGCAGT	20860
	GAGCTGGATG	TGCTGTATCC	GAGTGGTGCA	GGTCTGGCGT	GAGGTCCGCA	GGCCCTGCGG	TTAGCTGGAT	20930
	ATGCGGTGTC	GGATGGTGCA	GGTCCGGGTC	GAGGTACCCA	GGCCCTGCGG	TTAGCTGGAT	GTGCGGTGTC	21000
	CGGATGGTGC	AGGTCTGGGG	TGAGGTGCGC	AGGCCCTGCT	GTGAGCTGGA	TGTGCTGTAT	CCGGATGGTG	21070
75	CAGGTCCGGG	GTGAGGTGCG	CAGGCCCTG	GGTGAGCTGG	ATGTGCTGTA	TCCGGATGGT	GCAGGTCTGG	21140
	CGTGAGGTGC	CCAGGCCCTG	CGGTGAGCTG	GATGTGACAGT	GTACGGATGG	TGCAGGTCCG	GGGTGAGGTC	21210
	GCCAGGCCCT	GCGGTGGGCT	GTATGTGTGT	TGCTGGATG	GTGACAGGTC	GGGTGAGGTC	GCCAGGCCCT	21280

	TGCGGTGAGC	TGGATGTGTG	GTGCTCTGGAT	GCTGCAGGTC	CGGGGTGAGT	TCGCCAGGCC	CTCGGTGAGC	21350
	TGGATATGCG	GTGTCCCGCT	GTCCGAATGG	TGCAGGTCCA	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	21420
	GATGTGCCGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATGTGCCGT	21490
5	GTCCGGATGG	TGCAGGTCCG	GGGTGAGGTC	ACCAGGCCCT	CGGTGATCTG	GATGTGGCAT	GTCTTCTCG	21560
	TTTAAGGGGT	TGGCTGTGTT	CCGGCCCGCAG	AGCACCGTCT	CGGTGAGGAG	ATCTTGGCCA	AGTTCTCTGCA	21630
	CTGGCTGATG	AGTGTGTACG	TCGTGAGACT	GCTCAGGTCT	TTCTTTTATG	TCACGGAGAC	CACGTTTCAA	21700
	AAGAAACAGG	TCTTTTCTA	CCGGAAGAGT	GTCTGGAGCA	AGTTGCAAG	CATTGGAATC	AGGTACTGTA	21770
	TCGCCACGCC	AGGCCTCTGC	TTCTCGAAGT	CCTGGAACAC	CAGCCCGGCC	TACGATGCG	CCTGTCTCCA	21840
10	CTTGCCCTGTG	CTTCCCTGGC	TGTGCAGCTC	TGGGCTGGGA	GCCAGGGGCC	CCGTACACAG	CCTGGTCCAA	21910
	TGGATTCTG	TGCAAGGCTC	TGACTGCCTG	GAGCTCACGT	TCTCTTACTT	GTAATAACAG	GAGTTTGTGC	21980
	CAAGTGGTCT	CTAGGGTTTG	TAAAGCAGAA	GGGATTTAAA	TTAGATGGAA	ACACTACCAC	TAGCCTCCTT	22050
	GCCTTTCCCT	GGGATGTGGG	TCTGATTCTC	TCTCTCTTTT	TTTTTCTTTT	TTTGAGATGG	AGTCTCACTC	22120
	TGTTGCCAG	GCTGGAGTGC	AGTGGCATAA	TCTTGGCTCA	CTGCAACCTC	CACCTCCTGG	GTTTAAGCGA	22190
	TTACACAGCC	TCAGCCTCCT	AAGTAGCTGG	GATTACAGGC	ACCTGCCACC	ACGCCTGGCT	AATTTTGTGA	22260
15	CTTTTAGGAG	AGACGGGGTT	TCACCATGTT	GGCCAGGCTG	GTCTCGAACT	CATGACCTCA	GGTGATCCAC	22330
	CCACCTTGCC	CTCCCAAAGT	GCTGGGTTTA	CAGGCTAAGC	CACCGTGCCC	AGCCCCGAT	TCTCTTTTAA	22400
	TTTCATGCTGT	TCTGTATGAA	TCTTCAATCT	ATTGGATTGA	GGTCATGAGA	GGATAAAATC	CCACCCACTT	22470
	GGCGACTCAC	TGCAGGGAGC	ACCTGTGTCG	GGAGCACCTG	GGGATAGGAG	AGTTCCACCA	TGAGCTAACT	22540
20	TCTAGGTGGC	TGCATTGTAA	TGGCTGTGAG	ATTTTGTCTG	CAATGTTCCG	CTGATGAGAG	TGTGAGATTG	22610
	TGACAGATTG	AAGCTGGATT	TGCATCAGTG	AGGGACGGGA	GGCTGGTCT	GGGAGATGCC	AGCCTGGCTG	22680
	AGCCACGGCC	ATGGTATTAG	CTTCTCCGTG	TCCCGCCAG	GCTGACTGTG	GAGGGCTTTA	GTGCAAGATG	22750
	CAGGGCTTCC	CCAGCTCCCC	TGCACACTCG	AGTCCCTGGG	GGGCTTGTG	ACACCCCATG	CCCCAAATCA	22820
	GGATGTCTGC	AGAGGGAGCT	GGCAGCAGAC	CTGTCAGAG	GTAACACAGC	CTCTGGGCTG	GGGACCCCGA	22890
25	CGTGGTCTGC	GGGCCATTTC	CTTGATCTG	GGGGAGGGTC	AGGGCTTTCC	CTGTGGGAAC	AAGTTAATAC	22960
	ACAATGCACC	TTACTTAGAG	TTTACACGTA	TTTAATGGTG	TGCGACCCAA	CATGGTCATT	TGACCATGAT	23030
	TTTGGAAAGA	ATTAAATTGG	GGTGACCGGA	AGGAGCAGAC	AGACGTGGTG	GTCCCAAGA	TGCTCCTTGT	23100
	CACTACTGGG	ACTGTGTTC	TGCCTGGGGG	GCCTTGGAGG	CCCTCCTCC	CTGGACAGGG	TACCGTGCTT	23170
	TTTCTACTCT	GCTGGGCTG	CGGCCCTGGG	TCAGGCGACC	AGCTCCGGAG	CACCCGCGGC	CCCACTGTCC	23240
30	ACGGAGTGCC	AGGCTGTGAG	CCACAGATGC	CCAGTCCAG	GTGTGGCCGC	TCCAGCCCCC	GTGCCCCCAT	23310
	GGGTGGTTTT	GGGGGAAAG	GCCAAGGGCA	GAGGTGTCAG	GAGACTGGTG	GGCTCATGAG	AGCTGATTCT	23380
	GCTCCTTGGC	TGAGCTGCCC	TGAGCAGCCT	CTCCCGCCCT	CTCCATCTGA	AGGGATGTGG	CTCTTTCTAC	23450
	CTGGGGGTCC	TGCTTGGGGC	CAGCCTTGGG	CTACCCCACT	GGCTGTACCA	GAGGGACAGG	CATCCTGTGT	23520
	GGAGGGGCAT	GGGTTACAGT	GGCCCCAGAT	GCAGCCTGGG	ACCAGGCTCC	CTGGTGTGTA	TGGTGGGACA	23590
35	GTACACCTGG	GGGTTGACCG	CCGGACTGGG	CGTCCCCAGG	GTGACTATA	GGACAGGTG	TCCAGGTGCC	23660
	CTGCAAGTAG	AGGGGCTCTC	AGAGGGCTCT	GGCTGGCATG	GGTGGACGTG	GCCCCGGGCA	TGGCCTTCAG	23730
	CGTGTGCTGC	CGTGGGTGCC	CTGAGCCCTC	ACTGAGTCGG	TGGGGGCTTG	TGGCTTCCCG	TGAGCTTCCC	23800
	CCTAGTCTGT	TGTCTGGCTG	AGCAAGCCTC	CTGAGGGGCT	CTCTATTGCA	GACAGCACTT	GAAGAGGGTG	23870
40	CAGCTGCCGG	AGCTGTCCGA	AGCAGAGGTC	AGGCAGCATC	GGGAAGCCAG	GCCCCGCCCTG	CTGACGTCCA	23940
	GACTCCGCTT	CATCCCCAAG	CCTGACGGGC	TGCGGCCGAT	TGTGAACATG	GACTACGTCT	TGGGAGCCAG	24010
	AACGTTCCCG	AGAGAAAAGA	GGGTGGCTGT	GCTTTGGTTT	AACCTCCTTT	TTAAACAGAA	GTGCTTTTGA	24080
	GGCCACATT	TGGTATCAGC	TTAGATGAAG	GGCCCGGAGG	AGGGGCCACG	GGACACAGCC	AGGGCCATGG	24150
	CACGGCGCCA	ACCCATTGTG	CGGCACAGTG	AGGTGGCCGA	GGTGCCGGTG	CCTCCAGAAA	AGCAGCGTGG	24220
	GGGTCTAGGG	GGAGCTCCTG	GGGCAGGGAC	AGGCTCTGAG	GACCACAAGA	AGCAGCCGGG	CCAGGGCCTG	24290
45	GATGCAGCAC	GGCCCGAGGT	CCTGGATCCG	TGTCCTGCTG	TGGTGGCGAG	CCTCCGTGGC	CTTCCGCTTA	24360
	CGGGGGCCCG	GGACCAGGCC	ACGACTGCCA	GGAGCCCAAC	GGGCTCTGAG	GATCCTGGAG	CTTGCCCCAC	24430
	GGCTCCTGCA	CCCCACCCCT	TGGGCTGCGG	TGGCTGCGGT	GACCCCGTCA	TCTGAGGAGA	GTGTGGGGTG	24500
	AGGTGGACAG	AGGTGTGGCA	TGAGGATCCC	GTGTGCAACA	CACATGCGGC	CAGGAACCCG	TTTCAACACG	24570
50	GGTCTGAGGA	AGCTGGGAGG	GGTCTTAGGT	CCCGGGTCTG	GGTGGCTGGG	GACACTGGGG	AGGGGCTGCT	24640
	TCTCCCTGTG	TGCTTATGAG	TGGGGTGGGC	ACTTGGCCGG	ATCCACTTTC	CTGACTGTCT	CCCATGCTGT	24710
	CCCGGCCAGG	CCGAGCGTCT	CACCTCGAGG	GTGAAGGCAC	TGTTACAGCGT	GCTCAACTAC	GAGCGGGCGC	24780
	GGCGCCCCGG	CCTCCTGGGC	GCCTCTGTGC	TGGGCTGGGA	CGATATCCAC	AGGGCCTGGC	GCACCTTCGT	24850
	GCTGCGTGTG	CGGGCCCGAG	ACCCGCCCGC	TGAGCTGTAC	TTTGTCAAGG	TGGGTGCGGG	GGACCCCGCT	24920
55	GAGCAGCCCT	GCTGGACCTT	GGGAGTGGCT	GCCTGATTGG	CACCTCATGT	TGGGTGGAGG	AGGTACTCCT	24990
	GGGTGGGCGG	CAGGAGTGC	AGGTGACCCT	GTCACTGTGT	AGGACACACC	TGGCACCTAG	GGTGGAGGCC	25060
	TTACAGCCTT	CCTGCAGCAC	ATGGGGCCGA	CTGTGCACCC	TGACTGCCCC	GGCTCCTATT	CCCAAGGAGG	25130
	GTCCCACTGG	ATTCCAGTTT	CCGTACAGAA	AGGAACCCGA	ACGGCTCAGC	CACCAAGGCC	CGGTGCTTGT	25200
	CACCCCACTG	CTGAGCCAGG	GGTCTCCTGT	CCTGAGGCTC	AGAGAGGGGA	CACAGCCCGC	CCTGCCCTTG	25270
60	GGGTCTGGAG	TGGTGGGGGT	CAGAGAGAGA	GTGGGGGACA	CCGCCAGGCC	AGGCCCTGAG	GGCAGAGGTG	25340
	ATGTCTGAGT	TTCTGCGTGG	CCACTGTGAG	TCTCTCGCC	TCCACTCACA	CAGGTGGATG	TGACGGGGCG	25410
	GTACGACACC	ATCCCCCAGG	ACAGGCTCAC	GGAGGTGATC	GCCAGCATCA	TCAAAACCCA	GAACACGTAC	25480
	TGCGTGGCTC	GGTATGCCGT	GGTCCAGAAG	CGCCGCCATG	GGCACGTCCG	CAAGGCCCTT	AAGAGCCACG	25550
	TAAGGTTTAC	GTGTGATAGT	CGTGTCCAGG	ATGTGTGTCT	CTGGGATATG	AATGTGTCTA	GAATGCACTC	25620
65	GTGTCTGTGA	TGCGTTTCTG	TGGTGGAGGT	ACTTCCATGA	TTTACACATC	TGTGATATGC	GTGTGTGGCA	25690
	CGTGTGTGTC	GTGTGTGATG	TATCTGTGCG	GTGCATATTT	GTGGTGTGTG	TGTGTGTGGC	ACGTGTGTGT	25760
	CCATGGTGTG	TGTGCTGTG	GTGTGCTGTC	GTGTGTGTCT	GTGACACGTG	CATGTTTCATG	CTGTGTGCTG	25830
	CATGTCTGTG	ATGTGCTTAT	TTGTGGTGTG	TGTGTGATG	TGTCCGTGAC	ATATGCGTGT	CTATGGCATG	25900
	GGTGTGTGTG	GCCCTTGGGC	CTTACTCCTT	CCTCTCCAG	GCATGGTCCG	CACCATTTGC	CTCACGCTCT	25970
70	CGGTGTCTGG	TTTGGGGAGC	TCCACATTCA	GGGTCTCTAC	TTCTAGCATG	GGTGCCCTTG	TCCTGTGACA	26040
	GGGCTGGGCC	TTGAGACTG	TAAAGCAGGT	TTGAGAGGAG	AGTAGGGATG	CTGGTGGTAC	CTTCTGGGAC	26110
	CCCTGGGCACC	CCAGGACACC	CAGTCTGGCC	TATGCCGGCT	CCATGAGATC	TAGGAAGGCT	GATTTCAGGCC	26180
	TCGCTCCCGG	GGACACACTC	CTCCAGAGCC	GGCCGGGGGC	CTTGGGGCTG	GGCAGGGGTC	AAAGGGGCC	26250
	TGGGCTTGGG	TTCCCAACCA	GTGGTCATGA	GCACGCTGGA	GGGGTAAGCC	CTCAAGTCTG	TGCCAGGGCC	26320
	GGGTGCAGAG	GTGAAGAAGT	ATCCCTGGAG	CTTCCGTCTG	GGGAGAGGCA	CATGTGGAAA	CCCAAGGGA	26390
75	CCTCTTCTC	TGACTTCTTG	AGCT					26414



## Contig 2:

5 TGTGGGATTG GTTTTCATGT GTGGGATAGG TGGGGATCTG TGGGATTGGT TTTTATGAGT GGGGTAACAC 70  
 AGAGTTC AAG GCGAGCTTTC TTCCTGTAGT GGGTCTGCAG GTGCTCCAAC AGCTTTATTG AGGAGACCAT 140  
 ATCTTCTCTT GAACATATGGT CGGGTTTATA GTAAGTCAGG GGTGTGGAGG CCTCCCTGGG GCTCCCTGTT 210  
 CTGTTTCTTC CACTCTGGGG TCGTGTGGTG CCTGCTGTGG TGTGTGGCCG GTGGGAGGGG CTTCACAGGC 280  
 TCCTTGTGTT CATTGGCCTG GATGTGGCCC TGGCTACGCT CCGTCTTTGG AATTCCCTCG CGAGTTGGAG 350  
 GCTTCTTTTC TTTCTTTTTT TCTTCTTTTT TTTTTTTTTT TGATAACAGA GTCTCGCTCT TTTTGGCCCA 420  
 GGCTGGAGTG GTTTGGCGTG ATCTTGGCTC ACTGCAACCT GTGCTTCCTG AGTTC AAGCA ATTCTCTTGC 490  
 10 CTCAGCCTCC CAAGTAGCTG GAATTATAGG CGCCCACCAC CATGCTGACT AATTTTTGTA ATTTTAGTAG 560  
 AGACAGAGGT TCTCCATGTT GGCCAGGCTG GTCTCGAAGT CCTGACCTCA GGTGATCTCT CCACCTCGCG 630  
 CTCCCAAAGT GCTGGGATGA CAGGTGTGAA CCGCCGCGCC CGGCCGAGAC TCGCTTCTCG CAGCTTCCGT 700  
 GAGATCTGCA GCGATAGCTG CCTGCAGCCT TGGTGTGAC AACCTCCGTT TTCTTCTCC AGGTCTCGCT 770  
 AGGGGTCTTT CCATTTCATG ACTCTCTTCA CAGAAGAGTT TCACGTGTGC TGATTTCCCG GCTGTTCTCT 840  
 15 GCGTAATTGG TGTCTGCTGT TTATCGATGG CCTCTTCCA TTTCTTTAG GCTTTGTTTA TTGTGTTT 910  
 TCCGCTCCTT TGAAGGAAA GTTTCGATTA TGGATGTTG AACTTCTTT TCTAAACAAG CATCTGAAGT 980  
 TGGCGGTTTT CCTCTAAAGC AGGGATCCCG AGGCCCTGG CTGTGGAGTG GCACCGGTCT GGGGCTGTT 1050  
 AGGAACCCGG CGCACAGCGG GAGGCTAGGT GGGGTGTGGG GAGCCAGCGT TCCCGCTGA GCGCCGCCCC 1120  
 20 TCTCAGATCA GCGATGGCAT GCGGTGCTCA GAGGCGCACA CACCTACTG AGAAGTGTG GTGAGAGGGG 1190  
 TCTAGATTCT GTGCTCTTA TGGGAATCTA ATGCTGTATG ATCTGAGGTG GAACCGTTTG CTCCAAAAC 1260  
 CATCCCTTTC CCCACTGCTG TCCTGTGGAA AAATCGTCTT CCACGAAACC AGTCCCTGGT ACCACAATGG 1330  
 TTGGGGACCC TGTGCTAAAG ACCTGCTTCA GCAGCCTCTC GTCAGTGTG ATATATTGGC TTTCTGTGTT 1400  
 TGAGTCCAGA ATAATTACGG ATTTCTGTGA TGTCTTCCG CGACCTCAGA CCCATGGGCT ATTTGTGGGG 1470  
 25 GTGTGCGCTG CTCCTGGGTT GGG AAGGGTG CAGGCCCAT GTACCTTCTT GTTACTGCTT TCCAGGTGG 1540  
 TTCTCAGGGT TGAATCGTAC TCGATGTGGT TTTAGCCAC GGCCTGCGG CCAGCTCCTG GGGGCTGGGG 1610  
 AACATGCTGA AGCACAGAGT CACCGTGGCG GTCTTTTGT GCCTCACAAG CTCGAGGCGT CATGTGTCG 1680  
 TGTAGTGTG TGTACGTGC CTGCTCACAT CTGTCTTGG GGACGAGGG GCTTAGCAG TCCCGTAGTA 1750  
 AATGACAAGC GTCTGGGGG AGTCTGCAGA ATAGGAGGTG GGGGTGCGG TCTCTCTCCC GCGTCTTCAG 1820  
 ACTCTTCTCC TGCTGTGCT GTGGCTGCAC CTGCATCCCT GCAATCCCTC CAGCACTGGG CTGGAGAGGC 1890  
 30 CCGGAGCTC GAGTGCCACT TGTGCCACGT GACTGTGGAT GGCAGTGGT CACGGGGGTG TGAATGTGG 1960  
 TGACTGTGA TGGCGGTTGG TCACAGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2030  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2100  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2170  
 35 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2240  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2310  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2380  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2450  
 40 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2520  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2590  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2660  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2730  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2800  
 45 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2870  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 2940  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3010  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3080  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3150  
 50 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3220  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3290  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3360  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3430  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3500  
 55 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3570  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3640  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3710  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3780  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3850  
 60 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3920  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 3990  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4060  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4130  
 65 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4200  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4270  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4340  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4410  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4480  
 70 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4550  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4620  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4690  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4760  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4830  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4900  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 4970  
 TGTGGTGA CTGATGTGG GTGCTGGGGT CTGATGTGTG GTGACTGTGG ATGGCGGTG TGGGCTGA 5040

	CTGCCCGGGG	CCACCAGAGT	CTCCTTTTCT	GGCCCCGGCC	CCCTCCGGCT	CCTGGGCTGC	AGGCTCCCGA	5110
	GGCCCCGGAA	ACATGGCTCG	GCTTGGCGCA	GCCGGAGCGG	AGCAGGTGCC	ACACGAGGCC	TGGAAATGCG	5180
	AAGCCGGGTG	TGGAGTTGCT	CCTGCGTGGA	GGACGAGGGG	CGGGGGGTGT	GTCTGGGTCA	GGTGTGCGCC	5250
	GAGCGTTTGA	GCCTGCAGCT	TGTCAGCTCC	AAGTTACTAC	TGACGCTGGA	CACCCGGCTC	TCACACGCTC	5320
5	GTATCTCTCT	CTCCCCGATC	AAAAGGATTT	TATCCGATTC	TCATTCTGTG	CCCTGTCGTG	TGACCCCCGC	5390
	GAGGGCGCGG	GCTCTTCTCT	CTGTGACTAG	ATTTCCCATC	TGGAAGGTGC	GGGGTTGACC	GTGTAGTTTG	5460
	CTCCTCTCTG	GGGGCCTGTG	GTGGCCATGG	GGCAGGCGGC	CTGGGAGAGC	TGCCGTCAAC	CAGCCTCTGG	5530
	GTGAGCCACA	CTCACGGTGG	TAGAGCCACA	GTGGCTGGTG	CCACATCAGC	TCTCTCTGGT	TTTAAAGTAA	5600
	ACCACACACC	TCCCGGCGAG	CATCTGCCTG	CGACCCCTGT	TGTGCTCTGG	GAGAGTGGTA	GCACGGAGGA	5670
10	AATTCGTGCA	CACCTCAAGT	CATCAGCAAG	GTCAATCCGA	GTCAAGTGGG	ACGTGGAGGC	CTCTCTCTGG	5740
	GATCGTCTCC	AGCGGATAAA	GGACTGTGCA	CAGCTTCGGA	AGCTTTTATT	TAAAAATATA	ACTATTAAAT	5810
	ATTGCTATTAT	AAGTAATCAC	TAATGGTATC	AGCAATTATA	ATATTTATTA	AAGTATAAAT	AGAAATATTA	5880
	AGTAGTACAC	ACGTTCTGGA	AAAACACAAA	TGACATATGG	CAGCAGAGTG	AATTTTGGCC	GAGGGACACG	5950
	TGTGCACATG	TGTGTAAGCG	GCCCCAGGCC	CCACAGAAAT	CGCTGACAAA	GTCACTCTCC	CAGAGAAGCC	6020
15	ACCAAGGGCC	TCTTCTGTGG	TCGTGAATTT	TATTAAGATG	GATCAAGTCA	CGTACCGTCC	ACGTGTGGCA	6090
	GGGCTTTGGG	GAATGTGAGG	TGATGACTGC	GTCTCATATC	CCTGACAGAC	AGGAGGTGAC	TGTCTCTGTC	6160
	CTGTCTCTAG	GACACGGACA	GGCCCCAAGC	TCTAGTCCCC	ATCGTGGTCC	AGTTTGGCCT	GTGAATAAAA	6230
	ACGTCTTCAA	AACCTGTTGC	CCCAAAAACT	AAGAACAGAG	AGAGTTTCCC	ATCCCATGTG	CTCACAGGGG	6300
20	CGTATCTGCT	TGCGTTGACT	CGCTGGGCTG	GCCCGACTCC	TAGAGTTGGT	GCGTGTGCTT	CTGTGCAAAA	6370
	AGTGCACTCC	TCTTGCCCAT	CACGTGTGTA	TCTGCACCAG	CAAGGAAAGC	CTCTTTTCTT	TTCTTTCTTT	6440
	TTTTTTTTTT	GAGACGGAAC	GTCACTGTGG	TCTGCCTGGG	CTTGAGTGCA	GTGGCGCGAT	CTCAACTCAC	6510
	TGCAACCTCC	GCCTCCCGGG	TTCAGCAATT	TCTCTCTGCT	CAGCCTCCCG	AGCAGCTGAG	ATTACAGGCA	6580
	CCCCACCCCT	GCGCCTGGCT	AATTTTGTGA	TTTTTAGTAG	AGAGGGGTTT	TGCCCATGTT	GGCCAGGCTG	6650
25	GTCTCGAACT	CCTGACCTCA	GGTGATCCAC	CCACCTCGGC	CTCCCAAGT	GCTGGGATTA	CAGGTGTGAG	6720
	CCATCAGGCC	CAGCCGGAAA	GCCTCTTTTT	AAGGTGACCA	CCTATAGCGC	TCCCGAAAA	TAACAGGTCT	6790
	TGTTTTTGCA	GTAGGCTGCA	AGCGTCTCTT	AGCAACAGGA	GTGGCGTCTC	GTGGGCTCTG	GGGATGGCTG	6860
	AGGGTCGCGT	GGCAGCCATG	CCTTCTGTGT	GCACCTTTAG	GTTCACAGGG	GCTATTCTGC	TCTCACTGTT	6930
	TGTCTGAAAA	CGCACCCCTG	GCATCCTTGT	TGGAGAGATT	TCTGCTCTCT	GTTGGTCATG	CTGAAACTAG	7000
30	GGGCAAGGTT	GTATCCGTTG	GCGGCGAGCG	GCTACATGTA	GGGTGATGAG	TCTTTCACCG	TGGACAAATT	7070
	CCTTGAAAAA	AAAAAAGGGA	GTCCGGTTAA	GCATTCAATC	CGGGTCAAGT	GTCTGTTTCT	GTGAATAAAC	7140
	TCTAAGATT	AAGAAACCTT	AATGAAGAA	AACCTTGATG	ATTACAGACA	AGGATGTGGT	CACACCTGTG	7210
	GCTGGATCTG	TTTCAGCCGC	CCCAGTGCAT	GGTGAGAGTG	GGGAGCAGGG	ATTGTTTGTG	CAGAGGTCTC	7280
	ATCTGGTATG	TTTCTGAGGT	GTTTGCCGCG	TGAATGGTAG	ACGTGTCGTT	TGTGTGTATG	AGGTTCTGTG	7350
35	TCTGTGTGTG	GCTCGGTTTG	AGTGTACGCA	TGTCCAGCAC	ATGCCCTGCC	CGTCTCTCAC	CTGTGCTTCT	7420
	CCGCCCCAGG	TCCTACGTCC	AGTGCCAGGG	GATCCCGCAG	GGCTCCATCC	TCTCCACGCT	GCTCTGCAGC	7490
	CTGTGCTACG	GCGACATGGA	GAACAAGCTG	TTTGCGGGGA	TTGCGCGGGA	GCGGTGAGGC	CTCTCTTCTC	7560
	CCAGGGGGGG	TTGGGTGGGG	GTTGATTGCG	TTTTGATGCA	TTCACTGTGA	ATATTCCTGG	TGCTCTGGAG	7630
	ACCATGACTG	CTCTGTCTTG	AGGAACACGA	CAAGGTTGCA	GCCCTTCTTT	GGTATGAAGC	CGCACGGGAG	7700
40	GGGTGTCACA	GCCTGAGGAG	TGCGGGCTCC	ACGCAGGCTC	TGTCACGCGG	CCATGTCCAG	AGGCTCTCAG	7770
	GCTCAGCAGG	CGGGAGGGCC	GCTGCCCTGC	ATGATGAGCA	TGTGAATTCA	ACACCGAGGA	AGCACACACG	7840
	CTTCTGTAC	GTACCCAGG	TTCCGTTAGG	GTCTTGGGG	AGATGGGGCT	GGTGACGCTC	GAGGCCCCAC	7910
	ATCTCCACGC	AGGCCCTCGA	CAGGTGGGCT	GGACTGGGCG	CCTCTTCAGC	CCATTGCCCA	TCCCACTTGC	7980
45	ATGGGGTCTA	CACCCAAGGA	CGCACACACC	TAAATATCGT	GCCAACCTAA	TGTGGTTCAA	CTCAGCTGCG	8050
	TTTTATTGAC	AGCAGTTACT	TTTTTTTTTT	TAATACTTTA	AGTTCTAGGG	TACATGTGCA	CGACGTGCAG	8120
	GTTAGTTACA	TATGTATACA	TGTGCCATGT	TGGTGTGTGT	CACCCATTAA	CTCATCATTT	ACATTAGGTA	8190
	TATCTCTCAA	TGCTATCCCT	CCCCACTCCC	CCCATCCCAT	GACAGGCCCT	GGTGTGTGAT	GTTCGCCACC	8260
	CTGTGTCCAA	GTGTTCTCAT	TGTTCACTTC	CCACCTGTGA	GTGAGAACAT	GTGGTGTGTT	GTTTCTCTTC	8330
50	CTTGCAATAG	TTTGCTCAGA	GTGATGGTTT	CCAGCTTCGT	CCATGTCCCT	ACAAAGGACA	TGAACCTATC	8400
	CTTTTTTATG	ACTGCATAGT	ATTCGTTGGT	GTATATGTGC	CACATTTTCT	TAATCCAGTC	TATCATCGAT	8470
	GGACATTGCG	GTTGGTTGCA	AGTCTTTGCT	ACTGTGAATA	GTGCCCAAT	AAACATACGT	GTGCATGTGT	8540
	CTTTATAGCA	GCATGATTAA	TAATCCTTTG	GGTATATACC	CAGTAATGGG	ATGGCTGGGT	CAAAATGGTAT	8610
	TTCTAGTTCT	AGATCCTTGA	GGAATCACCA	CACGTCTCTC	CACAAATGGT	GAACATAGTT	ACACTCCAC	8680
55	CAACAGTGTA	AAAGTGTTC	GGTGCTGGAG	AGGATGTGGA	CAGCAGTTAT	TTTTTTATGA	AAATAGTATC	8750
	ACTGAACAAG	CAGACAGTTA	GTGAAGGATG	CGTCAGGAAG	CCTGCAGGCC	ACACAGCCAT	TTCTCTCGAA	8820
	GACTCCGGGT	TTTTCTGTG	CATCTTTTGA	AACTCTAGCT	CCAATTATAG	CATGTACAGT	GGATCAAGGT	8890
	TCTTCTTCAT	TAAGGTTCAA	GTTCTAGATT	GAAATAAGTT	TATGTAACAG	AAACAAAAAT	TTCTTGATCA	8960
	CACAACCTGC	TCTGGGATTT	GGAGGAAAGT	GTCTCTGAGC	TGGCGGCACA	CTGGTCAGCC	CTCTGGGACA	9030
	GGATACCTCT	GGCCCATGGT	CATGGGGGCG	TGGGCTTGGG	CCTGAGGGTC	ACACAGTGCA	CCATGCCACG	9100
60	CTTCTGTGG	ATAGGATCTG	GGTCTCGGAT	CATGCTGAGG	ACCAAGCTG	CCATGCTGGT	AAAGGGCACC	9170
	ACGTGGCTCA	GAGGGGCGCA	GGTTCCACGC	CCCAGCTTTC	TTACCGTCTT	CAGTTATTTT	TCCTTAAGAG	9240
	TCTGAGAAAT	GGGGCCGCGC	CTGATGGCCT	TCGTTCTGCT	TCAGCTGGCA	CAGAATTGCA	CAAGCTGATG	9310
	GTAAACACTG	AGTACTTATA	ATGAATGAGG	AATTTGCTGA	GCAGTTAACT	GTAGAGAGCT	CGTCTGTTGG	9380
	AAAGAAATTT	AAGTTTTC	TTTAAACGCT	TTGGAGAATG	TTACTTTTAT	TATGGCTGTG	TAAATTTGTT	9450
65	GACATTCACT	CCCTCGTAGA	CAGATACTAC	GTAAAAAGTG	TAAAGTTAAC	CTTGCTGTGT	ATTTTCCCTT	9520
	ATTTTAGGCT	GCTCCTGCGT	TTGGTGGATG	ATTCTTTGTT	GGTGACACCT	CACCTCACCC	ACGCGAAAAAC	9590
	CTTCTCTCAG	TGAGGCCCCG	GCGGTGTGTC	TGTGGGGACC	TCCACAGCCT	GTGGGCTTTG	CAGTTGAGCC	9660
	CCCCGTGTCC	TGCCCTGTGC	ACCGCAGCGT	TGCTCTGSCC	AAGTCCCTCT	TCTCTGCCGG	TGCTGGATCC	9730
	GCAAGAGCAG	AGGCGCTTGG	CCGTGCACCC	AGGCTTGGGG	GCGCAGGGGC	ACCTTCGGGA	GGAGTGGGTG	9800
70	ACCGTGCAGG	CCCTGTGCTC	GCAGAGACGC	ACCCAGGTTA	CACACGTGGT	GAGTGCAGGC	GGTGACCTGG	9870
	CTCCTGCTGC	TCTTTGGAAA	GTCAAGAGTG	GCGGCTCCTG	GGGCCCCAGT	GAGACCCCCA	GGAGCTGTGC	9940
	ACAGGGCCTG	CAGGGCCGAG	GCGGCAGCCT	CCTCCCCAGG	GTGCACCTGA	GCCTGCGGAG	AGCAGGAGCT	10010
	GCTGAGTGAG	CTGCCCCACA	GCGTTGCTGT	CGGTACGCTT	CCTGCGTGGG	GTTGTTTGGG	ATCGGTGGGA	10080
	GAATTTGGAT	TGCTGAGTGT	CTGCTGTCTT	GAACACCGGA	GATGGCTAGG	AGTGGGTTTC	AGAGTTGATT	10150
75	TTTGTAATC	AAACTAAAA	CAGGCACAGG	GGACCTGGCC	TCAGCACAGG	GGATTCTCCA	ATGTGGTCCC	10220
	CCTCAAGGGC	GCCCCACAGA	GCGGTTGGGC	TGTTTTTAAA	GTGCGATTG	ACGAGGGACG	AGAAACCTTG	10290
	AAAGCTGTAA	AGGGAACCTT	CAGAAAAATG	GGCCGCCAGG	GGTGGTTTCA	GGTGCTTTGC	TGGGCTGTGT	10360
	TTTGAAAAAC	CCATTGGGAC	CCGCCCTCCA	AGTCCACCTC	CCAGGTCCAC	CCTCCAGGGC	GCGCCCTGGG	10430

	TGGGGGTATG	CCTGGCGTTC	CTTGTGCCGC	AGCCCCGAGC	ACAGCAGGCT	GTGCACATTT	AAATCCACTA	10500
	AGATTCACTC	GGGGGGAGCC	CAGGTCCCAA	GCAACTGAGG	GCTCAGGAGT	CCTGAGGCTG	CTGAGGGGAC	10570
	AGAGCAGACG	GGGAACCGTG	CTTCTGTGTG	GCAAGTTCCT	GAGGGTGTCTG	GCCAGGGAGG	TGGCTCAGAG	10640
5	TGTATGTTGG	GGTCCCACCG	GGGGCAGAAC	TCTGTCTCTG	ATGAGTCGGC	AGCCATGTAA	CAGGAAGGGG	10710
	TGGCCACAGG	GAGCTGGGAA	TGCACACGGG	GAGCTGCCCA	GCTGGCCGAG	GTCCAGGGC	CAGGCCACAG	10780
	GAAGGGCAGG	GGGACGCCCG	GGGCCACAGC	AGAGGCCGCA	GGAAAGGGAAG	GGGATGCCCA	GGCCAGAGCA	10850
	GAGGCTTACG	GGCACAGGGG	GGCTCCCTGA	GCTGGGTGAG	CGAGGCTCAT	GACTCGGCGA	GGGAACCTCC	10920
	TTGACGTGAA	GCTGACGACT	GGTGTGCCCC	AGCTCACAGC	CCAGCCAGGT	CCCCGCCCTG	AGCAGGAATC	10990
10	CAGAACCCTC	CCCTTTGTCT	AAAGCACAGC	AGATGCCTTC	AGGGCATCTA	GGAGAAAACA	GGCAAAAGTCG	11060
	TTGAGAAACG	TCTTAAAGA	AGGTGGGATG	GTGGCAATTT	CTTGTCCAGA	TTTTAGTCTG	CCCCGGACCA	11130
	CAGATGAGTC	TATAACGGGA	TTGTGGTGTT	GCCATGGGGA	CACATGAGAT	GGACCATCAC	AGAGGCCACT	11200
	GGGGCTGCAC	CTCCCATCTG	AGTCTGTGCT	GTCCCGGGTC	CAGGCCAGGT	TCTTGATGTC	TCACCTACCT	11270
	GTCTGCCCG	GGAGACAGGG	AAAGCACCCC	GAAGTCTGGA	GCAGGGCTGG	GTCCAGGGTC	CTCAGAGCTC	11340
15	CTGCCAGGCC	CAGCACCTCT	CTCCAAATCA	CCACTTCTCT	GGGTTTTC	AAAGCATTTA	ACAAGGGTGT	11410
	CAGGTTACCT	CCTGGGTGAC	GGCCCCGCAT	CTGGGGCTG	ACATTGCCCC	TCTGCCTTAG	GACCCTGGTC	11480
	CGAGGTGTCC	CTGAGTATGG	CTGCGTGGTG	AACTTGCCGA	AGACAGTGGT	GAACCTCCCT	GTAGAAGACG	11550
	AGGCCCTGGG	TGGCACGGCT	TTTGTTCAGA	TGCCGGCCCA	CGGCCCTATT	CCCTGGTGGC	GGCTGCTGCT	11620
	GGATACCCCG	ACCTTGGAGG	TGCAGAGCGA	CTACTCCAGG	TGAGCGCACC	TGGCCGGGAG	TGGAGCCTGT	11690
20	GCCCCGCTGG	GGCAGGTGCT	GCTGCAGGGC	CGTTGCGTCC	ACCTCTGCTT	CCGTGTGGGG	CAGCGGACTG	11760
	CCAATCCCAA	AGGGTCAGAG	GGCCACAGGT	GCCCCCTGCT	CCATCTGGGG	CTGAGCAGAA	ATGCATCTTT	11830
	CTGTGGGAGT	GAGGTGCTC	ACAACGGGAG	CAGTTTTCTG	TGCTATTTTG	GTAAGAGGAA	ATGGTGCACC	11900
	AGACCTGGGT	GCACTGAGGT	GTCTTCAGAA	AGCAGTCTGG	ATCCGAACCC	AAGACGCCCG	GGCCCTGCTG	11970
	GGCGTGAGTC	TCTCAAACCC	GAACACAGGG	GGCCTGTGG	GCATGAGTCC	CTCTGAACCC	GAGACCTCTG	12040
25	GGCCCTGCTG	GGCGTGAGTC	TCTCCGAACC	CAGAGACTTC	AGGGCCCTTT	TGGGCGTGAG	TCTCTCCGCT	12110
	GTGAGCCCCA	CAGTCCAAGG	CTCATCCACA	GTCTACAGGA	TGCCATGAGT	TCATGATCAC	GTGTGACCCA	12180
	TCAGGGGACA	GGGCCATGTT	GTGGGGGGGG	TCTCTACAAA	ATTCTGGGGT	CTTGTTTCCC	CAGAGCCCCA	12250
	GAGCTCAAGG	CCCCGTCTCA	GGCTCAGACA	CAAATGAATT	GAAGATGGAC	ACAGATGCAG	AAATCTGTGC	12320
	TGTTTCTTTT	ATGAATAAAA	AGTATCAACA	TTCCAGGCAG	GGCAAGGTGG	CTCACACCTA	TAATCCCAGC	12390
30	ACTTTGGGAG	GCCGAGGTGG	GTGGATCACT	TGAGGGCCAGG	AGTTTGAGGC	CAACCTAACC	AACATAGTGA	12460
	AATTCCATTT	CTACTTAAAA	AATACAAAAA	TTAGCTTGGC	CTGGTGGCAC	ACGCCTGTAG	TCCCCTGTAT	12530
	GCGGGAGGCT	GAGGCAGGAG	AATCATTTGA	ACCCAGGAGG	CAGAGGTTGC	AGTGAGCCGA	GATCACACCA	12600
	CTGCATCTCA	GGCTGGGCAA	CAGAGTGAGA	CTTCATCTTA	AAAAAATAAA	AAAAAGTATC	AGCATTTCCA	12670
	AACCATAGTG	GACAGGTGTT	TTTTTATTTT	GTCTTCGAT	AATATTTACT	GGTGCTGTGC	TAGAGGCCGG	12740
35	AACTGGGGGT	GCCTTCCTCT	GAAAGGCACA	CCTTCATGGG	AAGAGAAATA	AGTGGTGAAT	GGTTGTTAAA	12810
	CCAGAGGTTT	AAACTGGGGT	CCTGTCTGTC	TGAGTTAAAC	GTCCAGATCT	GGACTTTGCC	TCTTTCCAGA	12880
	ATGCTCCCTG	GGGTTTGCTT	CATGGGGGAG	CAGCAGGTGT	GGACACCCCT	GTGATGGGGG	AGCAGCAGGT	12950
	GCAGACGCC	TCATGATGGG	GGAGTGGCAG	GTGCAGACAC	CCTTGTGCAT	GGTGCCCGAG	ATGTCCCTGT	13020
	TGCAGCTCCC	TCCCAACAAG	GATGCCGGTC	TCTGTGCTC	CCACAGTCC	CTGCTTCCCT	CTCAGAGCCT	13090
40	TACCTGGTCC	TGGCCTCCAC	TGGCTTTGTC	TGCATGATT	CCACATTTCC	TGGGCTCCCA	GCACCTCTTC	13160
	GCCTCTCCCA	GGCACCTCTG	CAGTGCTGGC	CATACCAGTC	AGCTGTGAAC	TGCTCACTGC	TTATTTTGCT	13230
	CCCCATGAAA	TGTATTTTTT	AGGACAGGCA	CCCTGGTTTC	CAGCCTCTGG	CACAGCATCA	GTGAATGTTA	13300
	TTGAAGGACA	AAGGACAGAG	AAACAAATCA	GGAAATGGG	TTCTCTCTAA	ACACATTGCA	AAGCCACAGA	13370
	GGCTAGTGCA	GGATGGGTGG	GCATCAGGTC	ATCAGATGTG	GGTCCAATGC	CAGAATATTC	TGTGCTCCCA	13440
45	AAGGCCACTT	GGTCAGAGTG	TGTGTCTGCA	GAGGTGGCTC	TAAAGGCTCA	GCAGTGGAGG	CAGTGGTTCG	13510
	CCATACTCAG	GGTGAACCTA	CATCCTCTGC	GTCTGAAGTA	TACAGCAGAG	GCTTGAAGGG	CACTGTGGAG	13580
	AAGAAAACAG	GCAAAATGAT	TAAAGAAAGT	GAAAAAGGAA	AAGTGGTAAG	ATGGGAATTT	TCTTGTCCAG	13650
	ATTTTAGTCT	CCCCAACACC	AGCTCAGATG	GTAGAATGTG	GTCAAGAACTG	ATGGACAGAA	CAATAGAACA	13720
	AAACGGAAGC	CCTATCTCTC	AGAAACGTGT	GTAAATGTGG	TATGTGGCAC	AGCTGATGGA	AAAGAGAGTG	13790
50	TGTGTGTAAT	TTTTTTTTCT	GAGAAAACCTG	ACTGGAAGCA	AATAAGTTGT	GTCTTTACAG	CATATACCAG	13860
	AGCAGATTCT	AGGTAGAAGA	GGAGACACAT	GCAAAACAAC	CCAGCAACAG	AAATAAACA	AAAGACTCAA	13930
	AGGGAAGGGA	GGTGAACGTT	CCCTGGTTTG	GTGTTGGGGA	AGGACACACA	GGGAGGCCGA	TGAAACCAAGT	14000
	GAGGCAACGG	GCATTGCTTT	CAGTGCAGAG	AAACTCAGCT	TGCCCTGAGC	ACAGTGAAAA	TGGCCATTCC	14070
	CTGGAGCGTT	TGTGCACGTG	ATTTATTTAA	GGCGCCCTGT	GAGGTCTCTG	ACATTATCTC	TCTCACTTTG	14140
55	TTCTCCTAAC	CACCTGAGAG	GTAGAGGAGG	AAAGGCTCCA	GGGGAGCAGC	CGCCCTTGGT	CACCCAGCTG	14210
	GCAAAAGGCA	TGCATGATTG	CAGCCTGGCC	TCTGTCTCCG	GGGCCCTTGC	TCTGCCCGAG	GACCCACAC	14280
	AAGTCAGACC	CATAGGCTCA	GGGTGAGCCG	GAGCCCAAGG	TCGTGTTGGG	GATGGCTGTG	AAAGAGAAAC	14350
	TGGACGTCTG	ATGCACACTT	GGGAAGGTCC	TACCAGCAGC	GTCAAAGAAA	TGCATGTGAA	ACTGACAGCG	14420
	AGACCCATCC	CTCAAAGAAA	CGCACGTGAA	ACTGATGGCG	AGACCTGTCC	CCATCCCTCA	TGCTGGCTCC	14490
60	TTTTCTGGGC	TTGCCAAGAG	CCAGCATCAG	GTGAGGCA	GCTGGAAAGA	CTTTTCTGGA	AAGCAGCTTG	14560
	TTTGCATGGA	AGTCCCTACA	ATGTCTGTG	TCTTCCAGT	AATTCCACTT	CTGAAGTGAC	CAGACATTAT	14630
	CACGGGTCTT	ATTACCATTT	TCCAGTGTTC	CAGGCAGGGG	GACTTGCCAC	AGCAAGTCAC	GAACCTGCCC	14700
	AAATACAGGG	CTAAGGAGAT	ATTATGCATC	ACAAAACCTG	CTCTGCCATT	AAACATTTTT	CAAAGAAATTT	14770
	TTGAAGAATG	TTTAATGGCA	CAAAACGTTT	ATTCAATGT	AGCAGTGTTT	AAAGCTGGAT	GTAAGAAGAAC	14840
65	ACACCCCAAG	AGCCTGCGGT	GAATGTCTAT	TGTGTTTCAT	TTTGACATG	GACATACATG	GGCAGTGAGT	14910
	GGTGGTGAGG	CCCTGGAGGA	CATCGGTGGG	ATGCCCTCAT	CCTGCCCTC	TGGAGACACC	ATGTGTGCCA	14980
	CGTGCACTCA	CTGGAGCCCT	GTTTAGCTGG	TGCCACCTGG	CTCTTCCATC	CCTGAGATTG	AAACACAGTG	15050
	AGATTCCCCA	CGCCCAACTC	AGTGTCTCTC	CACAAAAAAC	CTGAGTCACA	CCTGTGTTCA	CTCAGGGGAC	15120
	GCCCCGGAGC	CAGGGCTCCA	CAGTTTATTA	TGTGTTTTTG	GCTGAGTTAT	GTGCAGATCT	CATCAGGGCA	15190
70	GATGATGAGT	GCACAAACAC	GGCCGTGCGA	GGTTTGATA	CACCTCAACAT	CACCTAGCCAG	GTCTGGTGGG	15260
	AGTTTGGTCA	TGCAGAGTCT	GGATGGCATG	TAGCATTGGG	AGTCCATGGA	GTGAGCACCC	AGCCCTGCTG	15330
	GGCTGCAGCG	CATGCCCCAG	GCAGGACAAG	GAAGCGGGAG	GAAGGCAGGA	GGCTCTTTGG	AGCAAGCTTT	15400
	CGAGGAGGGG	GCTGGGTGTG	GCGCAGGCAC	CTGTGTCTGA	CATTCCCTCC	TGTGTCTCAG	ATCTGCCCGG	15470
	ACCTCCATCA	GAGCCAGTCT	CACCTTCAAC	CGCGGCTTCA	AGGCTGGGAG	GAACATGCGT	CGCAAACTCT	15540
75	TTGGGGTCTT	GCGGCTGAAG	TGTCACAGCG	TGTTTCTGGA	TTTGCAAGTG	AGCAGGCTGA	TGTCAGGAC	15610
	AGAGTTTACA	GTTTCAGGAG	TGTGTGCCCA	AGTATGTGTG	TGTGTGTGTG	CGCGGCTGCC	TGCAAGGCTG	15680
	ATGGTGACTG	GCTGCACGTA	AGAGTGCATA	TGTACGCATA	TACAGCTGAG	CACATACATG	TGTGCATGTT	15750
	TGTACATGAA	GGCATGGCAG	TGTGTGCACA	GGTGTGCAAG	GGCACAAGTG	TGTGCACATG	CGAATGCACA	15820

	CCTGACATGC	ATGTGTGTTT	GTGCACAGTC	GTGTGGGCAT	TCACGTGAGG	TGCATGCGTG	TGGGTGTGCA	15890
	GTGTGAGTAG	CATGTGTGCA	CATAACATGT	ATTGAGGGGT	CCTCGTGTTC	ACCCCGCTAG	GTCCTCAGCA	15960
	CCAGTGCCAC	TCCTTACAGG	ATGAGACGGG	GTCCACAGCC	TTGGTGGGCT	GAGGCTCTGA	AGCTGCAGCC	16030
5	CTGAGGGCAT	TGTCCCATCT	GGGCATCCGC	GTCCACTCCC	TCTCCTGTGG	GCTTCTGTGT	CCACTCCCCC	16100
	TCTCCTGTGG	GCATTATACAT	CCACTCCACT	CCCTCTCTCC	TGTGGGCATC	CGCGTCCACT	CCCCCTCTCT	16170
	GTGGGCATCT	CGGTCCACCT	CCCCTCTCTG	TGGGCATTTC	CGTCCACTCC	CTCTCCTGGT	TCCTTCTGTT	16240
	CTTGGCCGAG	CCTCGGGGGC	AGGCAGATGA	CACAGAGTCT	TGACTCGCCC	AGGGTGGTTC	GCAGTGGCCG	16310
	GGTGAGGGCC	AGGCCGGATT	TCACTGGGAA	GAGGGATAGT	TTCTTGTCAA	AATGTCTCTC	TTTCTGTGTC	16380
10	CATCTGAATG	GATGATAAAG	CAAAAAGTAA	AAACTTAAAA	TCCCAGAGAG	GTTTCTACCG	TTTCTCACTC	16450
	TTTCTTGCCG	ACTCTAGGTG	AACAGCCTCT	AGACGGTGTG	CACCAACATC	TACAAGATCC	TCCTGCTGCA	16520
	GGGCTACAGG	TGAGCCGCCA	CCAAGGGGTG	CAGGCCACGC	CTCCAGGGAC	CCTCCGGGCT	CTGCTCACCT	16590
	CTGACCCGGG	GCTTCACTTT	GGAACTCCTG	GGTTTTAGGG	GCAAGGAATG	TCTTACGTTT	TCAGTGGTGC	16660
	TGCTGCCGTG	GCACAGTTCT	GTTCCGCTGG	CTCTGTGCAA	AGCACCTGTT	CTCCATCTCT	GGGTAGTGGT	16730
15	AGGAGCCGGT	GTGGCCCGAG	GTGTCCCCAC	TGTGCCCTGT	CACCTGGCCG	GGGACGTCTA	GGAGGCCATC	16800
	CCAGGGCAGC	AGGGGCATGG	GGTAAAGAGA	TGTTTTATGG	GAGTCTTAGC	AGAGGAGGCT	GGGAAGGCTG	16870
	CTGAACAGTA	GATGGGAGAT	CAGATGCCCG	GAGGATTTCG	GGTCTCAGCA	AAGAGGGGCG	AGGTGGGTGC	16940
	AGGTGAGGGT	CGCTGGCCCC	ACCCCGGGGA	AGGTGCAGCA	GAGCTGTGGC	TCCCACACA	GCCCGGCCAG	17010
	CACCTGTGCT	CTGGGCATGG	CTGTGCTCTT	GGAACTGTCC	CTGTCTGGC	TGGTCAGGGG	GTGCCCTGCG	17080
20	CAAGAATCGA	CAACTTTATC	ACAGAGGGAA	GGGCCAATCT	GTGGAGGCCA	CAGGGCCAGC	TTCTGCTGCG	17150
	AGTCAGGGCA	GGTGTGGCA	CAAGCCTCGG	GGCTGTACCA	AAGGGCAGTC	GGGACCCACA	GGCCCGGGCC	17220
	TCCACCTCAA	CAGGCCTCCC	GAGCCACTGG	GAGCTGAATG	CCAGGAGGCC	GAAGCCCTCG	CCCCATGAGG	17290
	GCTGAGAAGG	AGTGTGAGCA	TTTGTGTTAC	CCAGGGCCGA	GGCTGCCGGA	ATTACCGTGC	ACACTTGATG	17360
	TGAAATGAGG	TCGTGCTCTA	TCGTGGAAAC	CCAGCAAGGG	CTCAGGGGAG	AGTTTTCTCA	TACAAGGTGC	17430
25	TACCATGAAA	ATGGTTTTTA	ACCCGAGTGC	TTGGCCCTTC	ATGCTCTGGC	AGGGAGGGCA	GAGCCACAGC	17500
	TGCATGTTAC	CGCCTTTGCA	CCAGCTCCAG	AGGCTTGGGA	CCAGGCTGTC	TCAGTTCAGC	GSTGCTCGCG	17570
	GCTCAGACCG	CCCTCCTCTC	TGGCTTCTCT	CTCTGCCCTA	AATCTTCCCT	CGTTTGCATC	TCCTTGACGC	17640
	GTGCCCTGGC	CCTCGTGCAA	GCTGCTTGAC	TCCTTTCCCG	AAACCTTTGG	GGTGTGCTGG	ATACAGGTGC	17710
	CACCTGAGGAC	TGGAGGTGTC	TGACACTGTG	GTGACCCCA	GGGTCCAGCT	GGCGTGCTTG	GGGCCCTCCT	17780
30	GGGCGCATGAT	GAGGTTCAGAG	GAGTTTTCCC	AGGTGAAAC	TCCTGGGAAA	CTCCAGGGC	CATGTGACCT	17850
	GCCACCTGCT	CCTCCCATAT	TCAGCTCAGT	CTTGCTCTCA	TTTCCCCACC	AGGGTCTCTA	GCTCCAGGGA	17920
	GCTCCCGTAG	AGGGCCTGGG	CTCAGGGCAG	GGCGGCTGAG	TTTCCCCACC	CATGTGGGGA	CCCTTGGGTA	17990
	GTCCGCTGAT	TGGGTAGCCC	TGAGGAGGCC	GAGATGCGAT	GGGCCACGGG	CCGTTTCAA	ACACAGATGC	18060
	AGGCACGTGG	AAGGCCCAGG	AATCCCTTTC	CCTCGAGGCA	GGAGTGGGAG	AACGGAGAGC	TGGGCCCCGA	18130
35	TTTACGGGCA	CCAGGCTGCG	AGTGGGCGAG	GCTGTGGTGG	TCCACGTGGC	GCTGGGGGCG	GGGTCTGATT	18200
	CAAAATCCGCT	GGGGCTCGGC	CTTCTTGCCC	CGTGCTGGCC	GGCCCTCCAC	ACGGGCTTGG	GGTGGACGCC	18270
	CCGACCTCTA	GCAGGTGGCT	ATTTCTCCCT	TTGGAAAGAGA	GGCCCTCACC	CATGCTAGGT	GTTCCTCTCC	18340
	TGGGTACAGGA	CGGTGGCCGT	GTGGCAACCC	CGGGACCTTA	GGCTTATTTA	TTTGTTTAAA	AACATTCTCG	18410
	GCCTGGCTTC	CGTTGTGCT	AAATGGGGAA	AAGACATCCC	ACCTCAGCAG	AGTTACTGAG	AGGCTGAAAC	18480
40	CGGGGTGCTG	GCTTGACTGG	TGTGATCTCA	GGTCATTCCA	GAAGTGCTC	AGGAAGTCAG	TGAGACCCAG	18550
	TACATGGGGG	GCTCAGGCAG	TGGGTGAGAT	GAGGTACACG	GGGGGCTCAG	GCAGTGGGTG	AGGCCAGGTA	18620
	CATATGGGGG	TCAGGCACGT	GGTGAGATGA	GGTACACGGG	GGGCTCAGGC	AGAGGGTCAG	ACCAGGTACA	18690
	CGGGGGCTCT	GATCACACGC	ACATATGAGC	ACATGTGCAC	ATGTGCTGTT	TCATGTTAGC	CAGGTCTGTT	18760
	CACACCTGCC	CCAAAGTCCC	AGGAAGCTGA	GAGGCCAAAG	ATGGAGGCTG	ACAGGGCTGG	CGCGTGGGCT	18830
45	CACACCTGTA	GTCCAGCAC	TTTGGGAGGC	CGAGGCGAGA	GGATCCCTTG	AGCCACAGGAG	TTTAAGACCA	18900
	GGCTGAGCAA	CATAGTAGAA	CCCCATCTCT	ATGAAAAATA	AAAACAAAAA	TTAGCTGAAC	ATGGTGGTGT	18970
	GGCCCTGTAG	TTCCARTACT	TGGGAGGCTG	AAGTGGGAGG	ATCACTTGAG	CCCAGGAGGT	GGAGGTGCA	19040
	GTGAGCTGAG	ATTGCAACAC	TGTACTGCAG	CCTGGGTGAC	AGAGTGAGAG	CCCATCTCAA	CAACACAAAA	19110
	GAAGACTGAC	AAATGCAATT	TCTTGGAAAG	AAACATTAG	TAGGAACCTA	ACCTACACAC	AGAAGCCAAG	19180
50	TCGGTGTCTC	GGTGTCACTG	AGATGAGATG	ATGGGTCTCT	ACACCATCAC	CCCAGACCCA	GGGTATTATG	19250
	ACCACAGGGG	CGGGTGGCTC	AGAAGGGATG	CCGAGGACGT	TGATATACGA	TGACATCAAG	GTTGTCTGAC	19320
	GAAGGGCAGG	ATTGATGATA	AGTACCTGCT	GGTACACAAG	GAACAATGGA	TAAACTGGAA	ACCTTAGAGG	19390
	CCCTCCCGGA	ACAGGGGCTA	ATCAGAAGCC	AGCATGGGGG	GCTGGCATCC	AGGATGGAGC	TGCTTCAGCC	19460
	TCCACATGCG	TGTTATATACA	GATGGTGCAC	AGAAACGCAG	TGTACTCTGT	CACACACAGA	CACGCAGCTA	19530
55	CTCGCACACA	CAAGCACACA	CACAGACATG	CATGATGCA	TCCGTGTGTG	TGCACCTGTG	CCCATGAGGA	19600
	AACCCATGCA	TGTGCATTCA	TGCACGCACA	CAGGCACCGG	TGGGCCCATG	CCACACCCA	CGAGCACCGT	19670
	CTGATTAGGA	GGCCTTTCTT	CTGACGCTGT	CGCCATCTCT	CTCAGGTTTC	ACGCATGTGT	GCTGCAGCTC	19740
	CCATTTTCAT	AGCAAGTTTG	GAAGAACCCC	ACATTTTTC	TGCGCGTCAT	CTCTGACACG	GCCTCCCTCT	19810
	GCTACTCCAT	CCTGAAAGCC	AAGAAGCGAG	GTATGTGCAG	GTGCTTGCCC	TCAGTGGCAG	CAGTGCCTGC	19880
60	CTGCTGGTGT	TAGTGTGTCA	GGAGACTGAG	TGAATCTGGG	CTTAGGAAGT	TCTTACCCCT	TTTCGCATCA	19950
	GGAACTGGTT	TAACCCAACC	ACTGTACGGC	TCGTCTGCCC	GCCTCTCTGT	GGGGTGAAGC	GAGCACCTGA	20020
	TGGAAGGGAC	AGGAGCTGTC	TGGGAGCTGC	CATCTCTCCC	ACCTTGCTCT	GCCTGGGGAA	CGCGTGGGGG	20090
	GCCTGGTCTC	TCCTGTTTTC	CCCATTGGTG	GATTTGGGGG	GCCTGGGCTC	TCCTGTTTTC	CTGTGGTGGG	20160
	GATTGGGCTG	TCTCCCGTCC	ATGGCACTTA	GGGCCCTTGT	GCAAAACCCAG	GCCAAGGGCT	TAGGAGGAGG	20230
65	CCAGGCCACG	GCTACCCAC	CCCTCTCAGG	AGCAGAGGCC	GCATATCACC	ACGACAGAGC	CCCGCGCCGT	20300
	CCTCTGCTTC	CCAGTCAACG	TCCTCTGCCC	CTGGACACTT	TGTCCAGCAT	CAGGGAGGTT	TCTGATCCGT	20370
	CTGAAATTC	AGCCATGTGC	AACCTGCGGT	CCTGAGCTTA	ACAGCTTCTA	CTTTCTGTTC	TTTCTGTGTT	20440
	TGGGAAATTT	CACCTGGAGA	AGCCGAAGAA	AACATTTCTG	TCGTGACTCC	TGCGGTGCTT	GGGTGGGGAC	20510
	AGCCAGAGAT	GGAGCCACCC	CGCAGACCGT	CGGGTGTGGG	CAGCTTTCCG	GTGTCTCCTG	GGAGGGGAGC	20580
70	TGGGCTGGGC	CTGTGACTCC	TCAGCCTCTG	TTTTCCCCCA	GGGATGTCCG	TGGGGGCCAA	GGGCGCCGCC	20650
	GGCCCTCTGC	CCTCCGAGGC	CGTGCACTGG	CTGTGCCACC	AAGCATTCTT	GCTCAAGCTG	ACTCGACACC	20720
	GTGTCACTTA	CGTGCCACTC	CTGGGTCAC	TCAGGACAGG	CAAGTGTGGG	TGGAGGCCAG	TGCGGGCCCC	20790
	ACCTGCCCCAG	GGGTCACTCT	TGAACGCCCT	GTGTGGGGCG	AGCAGCTCA	GATGCTGCTG	AAGTGCACAG	20860
	GCCCGCCGGC	CTGACCTTGG	GGGCTGGAG	CCACGCTGGC	AGCCCTATGT	GATTAAACGC	TGGTGTCCCC	20930
	AGGCCACGGA	GCCTGGCAGG	GTCCCAACT	TCTTGAACCC	CTGCTTCCCA	TCTCAGGGGC	GATGGCTCCC	21000
75	CACGCTTGGG	AGCCTTCTGA	CCCTTGACCT	GTGTCTCTCT	ACAGCCTCTT	CCCTGGCTGC	TGCCCTGAGC	21070
	TCCTGGGGTC	CTGAGCAAGT	TCTCTCCCG	CCCCGCGCGT	CCAGCGTCAC	TGGGCTGCTT	GTCTGCTCGC	21140
	CCCGGTGGAG	GGGTGTCTGT	CCCTTCACTG	AGGTTCACAC	CAGCCAGGGC	CACAGGTTGC	AGGCCCTGCC	21210

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55

```

TGCCCGGCCA CCCACACGTC CTAGGAGGGT TGGAGGATGC CACCTCTGGC CTCTTCTGGA ACGGAGTCTG 21280
ATTTTGGCCC CGCAGCCAG ACGCAGCTGA GTCCGAAGCT CCCGGGGACG ACCTGACTG CCCTGGAGGC 21350
CGCAGCCAAC CCCGCACTGC CCTCAGACTT CAAGACCATC CTGGACTGAT GGCCACCCGC CCACAGCCAG 21420
GCCGAGAGCA GACACCAGCA GCCCTGTCTC GCCGGGCTCT ACCTCCCAAG GAGGAGGGGG CGGCCACAC 21490
CCAGGCCCGC ACCGCTGGGA GTCTGAGGCC TGAGTGAGTG TTTGGCCGAG GCCTGCATGT CCGGCTGAAG 21560
GCTGAGTGTC CGGCTGAGGC CTGAGCGAGT GTCCAGCCAA GGGCTGAGTG TCCAGCACAC CTGCGCTCTT 21630
CACTTCCCA CAGGCTGGCG CTCGGCTCCA CCCAGGGGCC AGCTTTTCCT CACCAGGAGC CGGCTTCCA 21700
CTCCCCACAT AGGAATAGTC CATCCCCAGA TTCCCCATTG TTCACCCCTC GCCCTGCCCT CCTTTGCTTT 21770
CCACCCCCAC CATCCAGGTG GAGACCCCTG GAAGGACCTT GGGAGCTCTG GGAATTTGGA GTGACCAAG 21840
GTGTGCCCTG TACACAGGCG AGGACCCCTG ACCTGGATGG GGGTCCCTGT GGGTCAAATT GGGGGGAGGT 21910
GCTGTGGGAG TAAATACTG AATATATGAG TTTTTCAGTT TTTGAAAAAA TCTCATGTTT GAATCCTAAT 21980
GTGCACTGCA TAGACACCAC TGTATGCAAT TACAGAAACC TGTGAGTGAA CGGGGTGGTG GTCAGTGCGG 22050
GCCCATGGCC TGGCTGTGCA TTTACGGAAG TCTATGAGTG AATGGGGTTG TGGTCAGTGC GGGCCCATGG 22120
CCTGGCTGGG CCTGGGAGGT TTCTGATGCT GTGAGGCAGG AGGGGAAGGA GGGTAGGGGA TAGACAGTGG 22190
GAGCCCCCAC CCTGGAAGAC ATAACAGTAA GTCCAGGGCC GAAGGGCAGC AGGGATGCTG GGGGCCAGC 22260
TTGGGCGGCG GGGATGATGG AGGGCCTGGC CAGGGTGGCA GGGATGATGG GGGCCCCAGC TGGGGTGGCA 22330
GGGGTGATGG GGGGGGCTGG TCTGGGTGGC GGGGAAGATG GGGAGCCCTG GCTGGGCCCC TTCTCCCTCT 22400
GCCTCCCAAC TGCAGCCGTG GATCCGGATG TGCTTCCCTG GTGCACATCC TCTGGGCCAT CAGCTTTTAT 22470
GGAGGTGGGG GGCAGGGGCA TGACACCATC CTGTATAAAA TCCAGGATTC CTCTCTCTGA ACGCCCAAC 22540
TCAGGTTGAA AGTCACATTC CGCCTCTGGC CATCTCTTAA AGAGTAGACC AGGATTCTGA TCTCTGAAG 22610
GTGGGTAGGG TGGGCGAGTG GAGGGTGTGG ACACAGGAGG CTTCAGGGTG GGGCTGGTGA TGCTCTCTCA 22680
TCTCTTTATC ATCTCCAGT CTCTCTCTCT ATCTCTTAT CATCTCCCAG TCTCATCTGT TTCTCTTAA 22750
TCTCCAGTCT TCATCTGTCA TCCTCTTACC ATCTCCAGT CTCTCTCTT ATCTCTTAT CTCTAGTCT 22820
CATCCAGACT TACCTCCAG GCGGGGTGCC AGGCTCGCAG TGGAGCTGGA CATACGTCCT TCCTCAGGCA 22890
GAAGGAACAT GAAGGATTGG AGAGAACAGG AGGGGCGGCT CAGAGGGAGC CAGTCTTGGG GTGAAGAAAC 22960
AGCCCTCTCT CAGAAGTTGG CTTGGGCCAC ACGAAACCGA GGGCCCTGCG TGAGTGGGTC CAGAGCCTTC 23030
CAGCAGGTCC CTGCTGGGCG CTTATGGTAT GGGCGGCTCC TACTGAGTGC ACCTTGACA GGGCTTCTGG 23100
TTTGAGTGCA GCCCGGACGT GCCTGGTGTG GGGGTGGGGC CTTATGGCCA CTGGATATGG CGTCATTAT 23170
TGCTGTGCTC TCAGAGAATG TCTGAGTGAC CGAGCCTAAT GTGTATGGTG GGGCCCAAGT CACAGACTGT 23240
GTGCTAAATG CACTCTGGTG CTTGGAGCCC CCGTATAGGA GCTGTGAGGA AGGAGGGGCT CTTGGCAGCC 23310
GGCCTGGGGG CGCCTTTGCC CTGCAAACTG GAAGGGAGCG GCCCGGGGCG CCGTGGGCGG ACGAGCCTCA 23380
GTGAGAGGTT GGACAGAACA GGGCGGGGAC TTCCAGGAG CAGAGGCGCG TGCTCAGGCA CACCTGGGTT 23450
TGAATCACAG ACCAACAGGT CAGGCCATTG TTCAGCTATC CATCTTCTAC AAAGCTCCAG ATTCCTGTTT 23520
CTCCGGGTGT TTTTGTGTA AATTTTACTC AGGATTACTT ATATTTTGT CTAAAGTATT AGACCTTTAA 23590
AAAAGGTATT TGCTTTGATA TGGCTTAACT CACTAAGCAC CTACTTTATT TGTCTGTTTT TATTTATTAT 23660
TATTTATTAT ATTAGAGATG GTGCTACTCT TGTCACCCAG GTTGTAGTG CAGTGGCACA GTCATGGCTC 23730
GCTGTAGCCG CAACCCCCCA GGCTCAAGTG ATCCTCCGCG CTCAGCTTCC CAGAGTGCTG GGAATACAGG 23800
TGTGAGCCAC TGCCCTTGCC TGGCACTTTT AAAAACCACT ATGTAAGGTC AGGTCCAGTG GCTTCCACAC 23870
CTGTATCCCC AGTAGTTTGG GAAGCCGAGG CAGAAGGATT GTCTAGGCC AGGAGTTTGA GACCAGCATG 23940
GGTAACATAG GGAGACCCCA TCTCTACAAA AAATGCAAAA AGTTATCCGG GCGTGGGGTG CAGCATCTGT 24010
AGTCCAGCT GCTCGGAGG CTGAGTGGGA GGATCGCTTG AGCCCGGGAG GTCATGGCTC CAGTGGCTG 24080
TGATTGTACC ATCGCATCC AGCCTGGGCA ACAGAGTGAG ACCCTGTCTC AAAAAAAGAG AAAAAAAGAG 24150
AAGGAGAAGG AGAAGAGAAG AAGAAGGAAG AAGGAAGAG AAGAAGGAG AAGAAGGAG AAGAAGGAG 24220
AAGGAGGCT GCTAGTGCT AGGTAGACTG TCAATCTCA GAGCAAAATG AAAATAACAA AGTTTTAAAG 24290
GGAAAGAAAA ACCCCAGCTC TTTGGACTTC CTTAGGCCCTG AACTTCATCT CAAGCAGCTT CCTTCCACAG 24360
ACAAGCGTGT ATGGAGCGAG TGAGTTCAAA GCAGAAAGG AGGAGAAGCA GGCAGGGTG GAGGCTGTGG 24430
GTGACACCAG CCAGGACCCC TGAAAGGGAG TGGTTGTTTT CCTGCCTCAG CCCACGCTC CTGCGGTTCC 24500
TGCACCTGCT GTAACCGTCC ATGTTGGTGC CAGGTGCCCA CCGTGGGAAG ATGCTGTGCA GGGGGCTTGC 24570
CAAACTTTGG TGGGTTTCAG AAGCCCCAGG CACTTGTGGC AGGCACAATT ACAGCCCTC CCAAGAGTGC 24640
CCCACGCTCT TCTCTGAA CCTGTGAATG GTCAACCCG AAGGCAGAGG CTGGTGAAGG CTGCGAGTGG 24710
AATCACGGCT GCCAGTCAG CGATCTTAA GTCATCTCTG ATTATCTGGT GGGCTGATA TGCCACAG 24780
GGTCCCTAGA AGTGAGAGAG GGAGGCGAGG GAGAGTCAGA GAGGGGACGT GAGAAGGACC ACTGGCCACT 24850
GCTGGCTTTG AGATGGAGGA GGGGGTCCCC AGCCAAGGA TGGGGGCGAG CGCTCCATGC TGGAAAAGCA 24920
AGCAATCCTC CCGGCTCCTG AGGGCACACG GCCCTGCCCA CGCTCGATT TCAGGCCAGT GGGACCTGTT 24990
TCAGCTTTCC GGCCTCCAGA CTTGTAAGAT GATGCGTTTG TGTTCAGCCA CTAAGCTGCA GTGATTCTG 25060
ACAGCAGCAA ATGGAATAGC AGTACAGGGA AATGAATACA GGCAGAGTTC TCAGAGTGAC TCTCAGCCCA 25130
CCCCGGG
  
```

## 60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz  
 der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der  
 Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist  
 in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon				Intron				3'-Exon							
	Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N	C	A	G		G
10	Häufigk.(%)	70	60	80		100	100	95	70			80	100	100		60

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);

Intron 2: bp 12951-21566 (SEQ ID NO 5);

Intron 3: bp 21763-23851 (SEQ ID NO 6);

5 Intron 4: bp 24033-24719 (SEQ ID NO 7);

Intron 5: bp 24900-25393 (SEQ ID NO 8);

5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden

10 Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

15 Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15);

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position  
21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

## Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCCGGGTCGGCGTCCGGCTGGGGTTGAGGGCGGCCGGGGGAACAGCGACATGCGGAGAGCAGCGCAGG  
CGACTCAGGGCGCTTCCCCCGCAG

5

## Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTCGAGGGCCAGGCCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGCAGGCAGAGCC  
CTGGTCCTCCTGTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTCGAGTGGACACGGTGATCTCTGCC  
TCTGCTCTCCCTCCTGTCCAGTTTGCATAAACTTACGAGGTTACCTTTCACGTTTTGATGGACACGCGGTTTCCAGGCGC  
10 CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGGCGCGGCAGTGGAGCCGGGTTGCCGCAATGGGGAGAAAGTGTCTGGAAG  
CACAGACGCTCTGGCGAGGGTGCTGCAGGTTACCTATAATCCTCTTCGCAATTTCAAGGGTGGGAATGAGAGGTGGGGA  
CGAGAACCCCTCTTCTGGGGGTGGGAGGTAAGGGTTTTGCAGGTGCACGTGGTCAGCCAATATGCAGGTTTGTGTTTA  
AGATTTAATTGTGTGTTGACGGCCAGGTGCGGTGGCTCACGCCGTAATCCAGCACTTTGGGAAGCTGAGGCAGGTGGA  
TCACCTGAGGTGAGGAGTTTGTAGACCACTGACCAACATGGTGAAACCTATCTGTACTAAAAATACAAAATTAGCTG  
15 GGCATGGTGGTGTGCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGC  
TGCACTGAGCTGAGATTGTGCCATTGTACTCCAGCCTGGGCGACAAGAGTGAACTCTGTCTTTAAAAAAAAGTGT  
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCCATCTTTAGGTAT  
GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTGTTGGTGGTGTTCAGGGG  
ATGGTGTCTGCTGGGCCCTGCCGTGTCCCACTGTTTTTCTGGATTGTATGTTGAGGAACCTCCGCTCCAGCCCCCTTT  
20 TGGCTCCAGTGCTCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCGATTTACCCCCCTCCCACTCCCAAGAC  
ATGTAAGACTTCCGCCATGCAGACAAGGAGGGTGACCTTCTTGGGGCTCTTTTTTTCTTTTTTTCTTTTATGTTGGC  
AAAAGTCATATAACATGAGATTGGCACTCCTAACACCGTTTTCTGTGTACAGTGCAGAATTGCTAACTCGGCGGTGTTTA  
CAGCAGGTTGCTTGAATGCTGCGTCTTGCCTGACTGGAAGTCCCTACCCATCGAACGGCAGCTGCTCACACCTGCTGC  
GGCTCAGGTGGACACGCCAGTCAAGCTCATGCAACCCAGTTTGTCTTTTGTGCTCCAGCTTCCTTCGTTGAG  
25 GAGAGTTTGTGTTCTCTGATCAGGACTCTGCTGTCTGCTGTTCTCTGACTTCAGATGAGGTCAATCTGCCCTGG  
CTTATGCAGGGAGTGAGGCGTGGTCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGTTGCCCCAGGTGTCCCT  
GTCACGTGTAGGGTGAGTGAGGCGCGGCCCGGGTGTCCCTGTCCCGTGCAGCGTGATTGAGGTGTGGCCCCCGGGTGT  
CCCTGTACGTGTAGGGTGAGTGAGGCGCCATCCCGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCGGTGCCCGG  
GTGTCCCTGTCCCGTGCAGGGTGAGTGAGGCACTGTCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGCGGTCC  
30 CCGGGTGTCCCTCTCAGGTGTAGGGTGAGTGAGGCGCGGCCCCAGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCACC  
GTCCCTGGGTGTCCCTCCAGGTATAGGGTGAGTGAGGCACTGTCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGG  
CGCGGCCCGGGTGTCCCTCTCAGGTGCAGGGTGAGTGAGGCGCTGTCCCTGGGTGTCCCTGTCTCGTGTAGGGTGAGT  
GAGGCTCTGTCCCAGGTGTCTTGGCGTTTGCTCACTTGAGCTTGCTCCTGAATGTTTGTCTTTCTATAGCCACAGCT  
GCGCCGGTTGCCATTGCCTGGGTAGATGGTGCAGGCGCAGTGTGGTCCCAAGCCTATCTTTTCTGATGCTCGGCTCT  
35 TCTTGGTCACCTCTCCGTTCCATTTTGCTACGGGGACACGGGACTGCAGGCTCTCGCCTCCCGGTGCCAGGCACTGCAG  
CCACAGCTTCAGGTCCGCTTGCTCTGTTGGGCTGGCTGCTCACCAGTGCCCGCCACATGCATGCTGCCAATACTCC  
TCTCCAGCTTGTCTCATGCCAGGGTGGACTCTGGGCTGCCTGTGTCTGCTGCCAGTGTGCTGGAGACATCCAGAA  
AGGGTTCTGTGCTCCCTGAAGGAAAGCAAGTCACCCCAGCCCCCTCACTTGCTGCTTTTCTCCCAAGCTGCCCTCTGC  
TTGGCCCCCTTGGGTGGGTGGCAACGCTTGTACCTTATTCTGGGCACCTGCCGCTCATTGCTTAGGCTGGGCTCTGCT  
40 CCAGTCGCCCCCTCACATGGATTGACGTCCAGCCACAGGTTGGAGTGTCTCTGTCTCTCTGCTCTGAGACCCACGTG



GAGGGCCGGTGTCTCCGCCAGCCTTCGTGAGCTTCCCTCTTGGGTCTTAGTTTTGAATTTCACTGATTTACCTCTGACG  
TTTCTATCTCTCCATTGTATGCTTTTTCTTGGTTTATCTTTCAITTCCTTTTCTAGCTTCTTAGTTTAGTCATGCCCTTTC  
CCTCTAAGTGCTGCCCTTACCTGCACCTGTGTTTTGATGTGAAGTAATCTCAACATCAGCCACTTTCAAGTGTCTTAAA  
ATACTTCAAAGTGTTAATACTTCTTTAAGTATTCTTATTCTGTGATTTTTTTCTTTGTGCACGCTGTGTTTTGACGTGA  
5 AATCATTTTGATATCAGTGACTTTTAAGTATTCTTAGCTTATTCTGTGATTCTTTGAGCAGTGAGTTATTGAACACT  
GTTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTTAAGTTATCATTTTATTATTGATTTCTAACTCAGT  
TGTGTAGTGGTCTGTATAATAACCAATTATTGAAGTTTGGGAGCCTTGCTTTGTGATCTAGTGTGTGCATGGTTTCCAG  
AACTGTCCATTGTAAATTTGACATCCTGTCAATAGTGGGCATGCATGTTCACTATATCCAGCTTATTAAGGTCCAGTGCA  
AAGCTTCTGTCTCCTTCTAGATGCATGAAATTTCAAGAAGGAGGCCATAGTCCCTCACCTGGGGGATGGGTCTGTTCAAT  
10 TCTTCTCGTTTGGTAGCATTTATGTGAGGCATTGTTAGGTGCATGCACGTGGTAGAATTTTATCTTCTGATGAGTGAA  
TCTTTTGGAGACTTCTATGTCTCTAGTAATCTAGTAATTTCTTTTTTAAATTGCTCTTAGTACTGCCACACTGGGCTTCT  
TTTGATTAGTATTTTCTGCTGTGTCTGTTTTCTGCCCTTAATTTATATATATATATATATTTTTTTTTTTTTTGGAGACA  
GAGTCTTGGTCTGTGCGCCAGGGTGAGTGCAAGTGGTGTGATCACAGGTGAGTGAACCTTTTACCTTCTGGCCTGAGCCGT  
CCTCTCACCTCAGCCTCCTGAGTAGCTGGAACGCAGACACGCCGCTACACCTGGCTAATTTTTTAAATTTTTTCTGGA  
15 GACAGGGTCTTGCTGTGTTGCCAGGCTGGTCTCAAACCTCTTGACTCAAGGGATCCATCTACCTCGGCTTCCCAAAGTG  
CTGAATTACAGGCATGAGCCACCATGTCTGGCCTAATTTTCAACACTTTTATATTCTTATAGTGTGGGTATGTCTGTGTA  
ACAGCATGTAGGTGAATTTCCAATCCAGTCTGACAGTCGTTGTTTAACTGGATAACCTGATTATTTTCAATTTTTTGTGTC  
ACTAGAGACCCGCTGGTGCACTCTGATTCTCCACTTGCCCTGTGTCATGTCTCGTTCCCTGTTTCTCACCACCTCTTG  
GGTTGCCATGTGCGTTTCTGCGGAGTGTGTGATCCTCTCGTTGCCCTCCTGGTCACTGGGCAATTTGCTTTTATTCT  
20 CTTTGCTTAGTGTACCCCTGATCTTTTTATTGTGCTGTGTTGCTTTTGTATTGAGACAGTCTCACTCTGTCAACCA  
GGCTGGAGTGAATGGCACAATCTCGGCTCACTGCAACCTCTGCCCTCCTCGGTTCAAGCAGTTCTCATTCCTCAACCTCA  
TGAGTAGCTGGGATTACAGGCGCCACCACCGCTGGCTAATTTTTGTATTTTAGTAGAGATAGGCTTTCACCATGT  
TGGCCAGGCTGGTCTCAAACCTCTGACCTCAAGTATCTGCCCGCCTTGGCCTCCACAGTGCTGGGATTACAGGTGCAA  
GCCACCGTGCCCGGCATACCTTGATCTTTTAAATGAAGTCTGAAACATTGCTACCCTTGTCTGAGCAATAAGACCTT  
25 AGTGTATTTAGCTCTGGCCACCCCGAGCCTGTGTGCTGTTTTCCCTGCTGACTTAGTTCTATCTCAGGCATCTTGACA  
CCCCCACAAGCTAAGCATTATTAATATTGTTTTCCGTGTTGAGTGTCTCTGTAGCTTTGCCCCCGCCCTGCTTTTCTCTC  
TTTGTTCCTGCTCTGCTCTCAGGCCCGCCGTCTGGGGTCCCTTCTTGTCTTTGCGTGGTTCTTCTGTCTTG  
TTATTGCTGGTAAACCCAGCTTTACCTGTGCTGGCCTCCATGGCATCTAGCGACGTCCGGGGACCTCTGCTTATGATGC  
ACAGATGAAGATGTGGAGACTCAGAGGAGGGCGGTCACTTGGCCCGTGAGTGTCTGGAGCACCAGTGCCAGCGTTC  
30 CTTAGCCAGTGAGTGACAGCAACGTCCGCTCGGCTGGGTTGAGCCTGGAACCCAGGCATGTGCGGGTCTGGTGGCT  
CCGCGGTGTGAGTTTGAAATCGCGCAACCTGCGGTGTGGCGCCAGCTCTGACGGTGTGCTGCGGGGGAGTGTCTG  
CTTCTCCCTTCTGCTTGGGAACCAGGACAAAGGATGAGGCTCCGAGCCGTTGTGCGCCAACAGGAGCATGACGTGAGCC  
ATGTGGATAATTTTAAATTTCTAGGCTGGGCGCGGTGGCTCACGCCCTGTAATCCAGCACTTTGGGAGGCCAAGGCGGG  
TGGATCACGAGGTGAGGAGTCTGAGACCATCTGGCCAAATGATGAAACCCCATCTGTACTAAAAACACAAAAATTAGC  
35 TGGGCGTGGTGGCGGGTGCTGTAAATCCAGCTACTCGGGAGGCTGAGGAGGAGAAATGCTTGAACCTGGGAGTTGGAA  
GTTGAGTGAGCCGACATTGCACCACTGCACTCCAGCCTGGCAACACAGCGAGACTCTGTCTCAAAAAAAAAAAAAA  
AAAAAAAAAAAAATCTAGTAGCCACATTAAAAAAGTAAAAAGAAAAGGTGAAATTAATGTAATAATAGATTTTACTGAA  
GCCCAGCATGTCCACACCTCATTTTTAGGGTGTATTGTTGGTGGGAGCATCACTCACAGGACATTTGACATTTTTTGAGC  
TTTGTCTGCGGGATCCCGTGTGTAGTCCCGTGTGGCCATCTCGGCTGGACCTGCTGGGCTTCCATGGCCATGGCT  
40 GTTGTACCAGATGGTGCAGGTCCGGGATGAGGTGCGCCAGGCCCTCAGTGAGCTGGATGTGAGTGTCCGGATGGTGCAGG  
TCTGGGATGAGGTGCGCCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTGAGGGTGAGGTCTCCAG

GCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGAGGTCCGGGGTGAGGTCCGACGGCCCTGCTGTGAGCTGGATG  
TGTGGTGTCTGGATGGTGCAGGTACGGGGTGAGGTCTCCAGGCCCTCGGTAAGCTGGAGGTATGGAGTCCGGATGATGCA  
GGTCCGGGGTGAGGTCCGACGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGGGTGAGGTCAAC  
AGGCCCTGCGGTGAGCTGGGTGTGCGGTGTCTGGATGGTGCAGGTCTGGAGTGGGTCCGACAGCGGTGCCAGACCATGC  
5 GGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTGCCAGGCCCTGCTGTGAGTTGGATGTGGGGT  
GTCCGGATGCTGCAGGTCCGGTGTGAGGTCAACAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCT  
GGGGTGAAGGTCCGACGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGAGTGGGTCCGAC  
GCCCTCGGTGAGCTGGATGTGAGTGTCCAGATGGTGCAGGTCCGGGGTGAGGTCCGACAGCCCTGCGGTGAGCTGGATG  
TGCGGTGTCTGGATGGTGCAGGTCTGGAGTGGGTCCGACGGCCCTCGGTGAGCTGGATGTATGGAGTCCGGATGGTGCC  
10 GGTCCGGGGTGAGGTCCGACAGCCCTGCTGTGAGCTGGATGTGCGGTGTCTGGATGGTACAGGTCTGGAGTGGGTCCGAC  
AGACCTGCTGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTACGGGGTGAGGTCTCCAGGCCCTCGGTGAGCTGGA  
GGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCGACGGCCCTGCTGTGAAGTGGATGTGCGGCTGTGGATGGT  
GCAGGTCTGGGGTGTGGTCCGACGGCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCC  
CCAGGCCCTGCTGTGAGCTGGATGTGCGGCTGTGGATGGTGCAGGTCTGGGGTGTGGTCCGACGGCCCTCGGTGAGCTG  
15 GAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATG  
GTGCAGTCCGGGGTGAGGTCCGACGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCTGGGGTGAGGT  
CACCAGGCCCTGCGGTGAGCTGGTGTGCGGTGTCCGGTGTGCTGCAGGTCCGGGGTGAGTCCGACAGGCCCTCGGTGAGC  
TGGATGTGCGGTGTCCCGTGTCCGGATGGTGCAGGTCCAGGGTGAGGTCCCTAGGCCCTTGGTGGGCTGGATGTGCCGT  
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGACGGCCCTTGGTGAGCTGGATGTGCGGTGTCTGCATGGTGCAGGTCTG  
20 GGGTGAGGTCCGACGGCCCTTGGTGGGCTGGATGTGTGGTGTCCGGATGGTGCAGGTCCGGCGTGAGGTCCGACGGCCCT  
GCTGTGAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTAGCCAAGGCCCTTGGTGAGCTGGATGTGGG  
GTGTCCGGATGGTGCAGGTCCGGGGTGAGGTCCGACGGCCCTGCGGTAGCTGGATATGCGGTGTCCGGATGGTGCAGGT  
CCGGGGTGAGGTCAACAGGCCCTGCGGTAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTCCGACGG  
CCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCAGTGGAGCTGGATG  
25 TGCTGTATCCGGATGGTGCAGGTCTGGCGTGAGGTCCGACGGCCCTGCGGTAGCTGGATATGCGGTGTCCGGATGGTGCA  
GGTCCGGGGTGAGGTCAACAGGCCCTGCGGTAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGAC  
AGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCGGTGAGCTGG  
ATGTGCTGTATCCGGATGGTGCAGGTCTGGCGTGAGGTCCGACGGCCCTGCGGTGAGCTGGATGTGAGTGTACGGATGG  
TGCAGGTCCGGGGTGAGGTCCGACGGCCCTGCGGTGGGCTGTATGTGTGTGTCTGGATGGTGCAGGTCCGGGGTGAGTT  
30 CGCCAGGCCCTGCGGTGAGCTGGATGTGTGGTGTCTGGATGTGCAGGTCCGGGGTGAGTTCCGACAGGCCCTCGGTGAGC  
TGGATATGCGGTGTCCCGTGTCCGAATGGTGCAGGTCCAGGGTGAGGTCCGACAGGCCCTTGGTGGGCTGGATGTGCCGT  
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGACGGCCCTTGGTGAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCCG  
GGGTGAGGTCAACAGGCCCTCGGTGATCTGGATGTGGCATGTCTTCTCGTTTAAG

35 Intron 3 (SEQ ID NO 6)

GTACTGTATCCCAAGCCAGGCCCTGCTTCTCGAAGTCTGGAACACAGCCCGGCCCTCAGCATGCGCCTGTCTCCACT  
TGCCTGTGCTTCCCTGGCTGTGAGCTCTGGGCTGGGAGCCAGGGGCCCGTCACAGGCCCTGGTCCAAGTGGATTCTGTG  
CAAGGCTCTGACTGCTGGAGCTCACGTTCTTACTTGTAAAATCAGGAGTTTGTGCCAAGTGGTCTCTAGGGTTTGT  
AAGCAGAAGGGATTAAATTAGATGGAAACATAACCACTAGCCCTTTCCTTTCCTTGGGATGTGGGTCTGATTCTCTC  
40 TCTCTTTTTTTTTCTTTTTTGGAGATGGAGTCTCACTCTGTTGCCAGGCTGGAGTGCAGTGGCATAATCTTGGCTCACT

GCAACCTCCACCTCCTGGGTTTAAGCGATTACCAGCCTCAGCCTCCTAAGTAGCTGGGATTACAGGCACCTGCCACCAC  
GCCTGGCTAATTTTGTACTTTTAGGAGAGACGGGGTTTACCATTGTTGGCCAGGCTGGTCTCGAACTCATGACCTCAGG  
TGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGTTTACAGGCTAAGCCACCGTGCCAGCCCCGATTCTCTTTAATT  
CATGCTGTCTGTATGAATCTTCAATCTATTGGATTAGGTATGAGAGGATAAAATCCACCCACTTGGCGACTCACTG  
5 CAGGGAGCACCTGTGCAGGGAGCACCTGGGGATAGGAGAGTTCCACCATGAGCTAACTTCTAGGTGGCTGCATTGGAATG  
GCTGTGAGATTTTGTCTGCAATGTTTCGGCTGATGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTGTCATCAGTGAG  
GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTTCTCCGTGTCCCGCCAGGC  
TGACTGTGGAGGGCTTTAGTCAGAAGATCAGGGCTTCCCCAGCTCCCTGCACACTCGAGTCCCTGGGGGGCCTTGTGAC  
ACCCCATGCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG  
10 GACCCGACGTGGTGTGGGGCCATTTCCTTGCATCTGGGGAGGGTCAGGGCTTTCCTGTGGGAACAAGTTAATACAC  
AATGCACCTTACTTAGACTTTACAGTATTTAATGGTGTGCGACCCAAACATGGTCATTTGACCAGTATTTTGGAAAGAAT  
TTAATTGGGGTGACCGAAGGAGCAGACAGACGTGGTGGTCCCCAAGATGCTCCTTGTCACTACTGGACTGTGTCTG  
CCTGGGGGGCCTTGGAGGCCCTCCTCCCTGGACAGGGTACCGTGCCTTTTCTACTGTCTGGGCTCGCGCCTGCGGTG  
AGGGCACCAGCTCCGGAGCACCCGCGGCCCACTGTCCACGGAGTGCCAGGCTGTGAGCCACAGATGCCAGGTCCAGGT  
15 GTGGCCGCTCCAGCCCCGTGCCCCATGGGTGGTTTGGGGGAAAAGGCCAAGGGCAGAGGTGTGAGGAGCTGGTGGG  
CTCATGAGAGCTGATTCTGCTCCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCCGCCTCTCCATCTGAAGGATGTGGCT  
CTTTCTACCTGGGGTCTGCTCTGGGGCCAGCCTTGGGCTACCCAGTGGCTGTACAGAGGGACAGGCATCCTGTGTGG  
AGGGGCATGGGTTACGTGGGCCAGATGCAGCCTGGGACCAGGCTCCCTGGTGTGATGGTGGGACAGTCAACCTGGGG  
GTTGACCGCCGACTGGGCGTCCCCAGGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAG  
20 AGGCGTCTGGCTGGCATGGGTGGACGTGGCCCCGGGCATGGCCTTCAGCGTGTGCTGCCGTGGGTGCCCTGAGCCCTCAC  
TGAGTCGGTGGGGCTTGTGGCTTCCCGTGAAGCTTCCCCCTAGTCTGTGTCTGGCTGAGCAAGCCTCCTGAGGGGCTCT  
CTATTGCAG

#### Intron 4 (SEQ ID NO 7)

25 GTGGCTGTGCTTTGGTTTAACTTCCTTTTAAACAGAAAGTGCCTTTGAGCCCCACATTTGGTATCAGCTTAGATGAAGGG  
CCCGGAGGAGGGGCCACGGGACACAGCCAGGGCCATGGCACGGCGCCAAACCATTTGTGCGCACAGTGAGGTGGCCGAGG  
TGCCGGTGCTCCAGAAAAGCAGCGTGGGGGTGTAGGGGGAGCTCCTGGGGCAGGGACAGGCTCTGAGGACCACAAGAAG  
CAGCCGGGCCAGGGCCTGGATGCAGCACGGCCCCAGGTCTCGATCCGTGTCTGTGTTGGTGCAGCCTCCGTGCGCT  
TCCGCTTACGGGGCCCCGGGGACAGGCCAGACTGCCAGGAGCCACCGGGCTCTGAGGATCCTGGACCTTGGCCACGG  
30 CTCTGCACCCACCCCTGTGGCTGCGGTGGCTGCGGTGACCCCGTCATCTGAGGAGAGTGTGGGTGAGGTGGACAGAG  
GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGGAACCCGTTTCAAACAGGGTCTGAGGAAGCTGGGAGGGG  
TTCTAGGTCCCGGTCTGGGTGGCTGGGGACACTGGGGAGGGGCTGCTTCTCCCTGGGTCCCTATGGTGGGGTGGGCAC  
TTGGCCGATCCACTTTCCTGACTGTCTCCATGCTGTCCCCGCCAG

#### 35 Intron 5 (SEQ ID NO 8)

GTGGGTGCCGGGACCCCCGTGAGCAGCCCTGCTGGACCTTGGGAGTGGCTGCCTGATTGGCACCTCATGTTGGGTGGAG  
GAGGTACTCCTGGGTGGGCGCAGGGAGTGCAAGTGACCTGTCACTGTTGAGGACACACCTGGCACCTAGGTTGGAGGC  
CTTCAGCCTTTCTGCAGCACATGGGGCCGACTGTGCACCTGACTGCCCGGGCTCCTATCCCAAGGAGGGTCCCACTG  
GATTCAGTTTCCGTGAGAGAAGGAACCGCAACGGCTCAGCCACCAGGCCCGGTGCCTTGACCCCACTGCTGAGCCAG  
40 GGGTCTCCTGTCTGAGGCTCAGAGAGGGGACACAGCCCGCTGCCCTTGGGGTCTGGAGTGGTGGGGTCTCAGAGAGAG

5'-Bereich Intron 6 (SEO ID NO 9)

15

TGTCGGGATTGGTTTTTCATGTGTGGGATAGGTTGGGGATCTGTGGGATTGGTTTTTATAGTGGGGTAACACAGAGTTCAGG  
GCGAGCTTTCTTCTGTAGTGGGTCTGCAGGTGCTCCAACAGCTTTATTGAGGAGACCATATCTTCTTTGAACTATGGT  
CGGGTTTATAGTAAGTCAGGGGTGTGGAGGCCTCCCTGGGCTCCCTGTTCTGTCTTCTTCCACTCTGGGGTCGTGTGGT  
CCTGCTGTGGTGTGTGGCCGGTGGGCAGGGCTTCCAGGCCTCCTTGTTGTTTCATTGGCCTGGATGTGGCCCTGGCTACGCT  
CCGTCTTGGGAATCCCTGCGAGTTGGAGGCTTTCTTCTTTCTTTTTTCTTCTTTTTTTTTTTTGGATAACAGA  
GTCTCGCTCTTTTTTGCCAGGCTGGAGTGGTTGGCGTGATCTTGGCTCACTGCAACCTGTGCTTCTCTGAGTTCAAGCA  
ATTCTCTTGCTCAGCCTCCCAAGTAGCTGGAATTATAGGCGCCACCACCATGTGACTAATTTTTGTAAATTTAGTAG  
AGACGAGGTTTCTCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCTCCACCTCGGCCCTCCCAAAGT  
GCTGGGATGACAGGTGTGAACCGCCGCGCCGGCCGAGACTCGTCTCTGCAGCTTCCGTGAGATCTGCAGCGATAGCTG  
CCTGCAGCCTTGGTGTGACAACTCCGTTTTCTTCTCAGGTCTCGCTAGGGGTCTTCCATTTCTGACTCTCTTCA  
CAGAAGAGTTTACGTTGTGCTGATTCCCGGCTGTTTCTCGCTAATTGGTGTCTGCTGTTTATCGATGGCCTCCTTCCA  
TTTCTTTTAGGCTTTGTTTATGTGTTTTTCCGGCTCCTTGAAGGAAAAGTTTCGATTATGGATGTTTGAACCTTTCTT  
TCTAACAAGCATCTGAAGTTGCCGTTTTCCCTCTAAGCAGGGATCCCGAGGCCCTGGCTGTGGAGTGGCACC GGTTCT  
GGGGCTGTTAGGAACCCGGCGCACAGCGGGAGGCTAGGTGGGGTGTGGGGAGCCAGCGTTCCCGCCTGAGCCCCGCCCC  
TCTCAGATCAGCAGTGGCATGCGGTGCTCAGAGGCGCACACCCCTACTGAGAACTGTGCGTGAGAGGGGTCTAGATTCT  
GTGCTCCTTATGGGAATCTAATGCCTGATGATCTGAGGTGGAACCGTTTGTCTCCAAAACCATCCCCTTCCCCACTGTCT  
TCTGTGGAAAAATCGTCTTCCACGAAACCAAGTCCCTGGTACCACAATGGTTGGGGACCCTGTGCTAAAGACCTGCTTCA  
GCAGCCTCTCGTCAGTGTGATATATGGCTTTTCTGTGTTGAGTCCAGAATAATTACGGATTCTGTGATGCTTTCCG  
CGACCTCAGACCCATGGGCTAATTGTGGGCGTGTGCCTGCTCCTGGGTGGGAAGGGTGCAGGCCCATGTACCTTCT  
GTTACTGCCTTCCAGGTTGGTTCTCAGGTTGAATCGTACTCGATGTGTTTTAGCCACAGGCCCTGCGCCAGCTCCT  
GGGGCTGGGGAACATGCTGAAGCACAGAGTACCCTGCGCGTCTTTTGATGCCTACAAGCTCGAGGCCCTCTGTGTCC  
TGTTAGTGTGTGTCAGTGCCTGCTCACATCCTGTCTTGGGGACGCAGGGGCTTAGCAGGTCCCGTAGTAAATGACAAG  
GTCCTGGGGGAGTCTGCAGAATAGGAGGTGGGGGTGCGGCTCTCTCTCCCGCTCTTTCAGACTCTTCTCCTGCGCTGTGC

40

GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGCCCGGAGCTCGAGTGCCACTTGTGCCACGT  
GACTGTGGATGGCAGTCGGTCACGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTTGGTCACAGGGGTCTGATGTGTG  
GTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGG  
ATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTG  
5 GGGTCTGATGTGTGGTGAAGTGTGGATGGCAGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATG  
TGGTGAAGTGTGGATGGCAGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGT  
GTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGG  
CGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGTGAATCGGTCA  
CAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGTGAATCGGTCA  
10 GGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTTGGTCCCGGGGG  
TCTGATGTGTGGTGAAGTGTGGATGGCGATCGGTCAAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCT  
GATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGT  
GACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGAT  
GGCGGTTGGTCCCGGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCAG  
15 TCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGG  
TCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGT  
GGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGTGAATCGGTCAAGGGGTCTGATGTGTGGT  
GACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGAT  
GGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTAGGGTCTGATGTGTGGTGAAGTGTGGATGGCAGTCG  
20 GTCACAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGG  
GGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGAT  
GTGGTGAAGTGTGGATGGTGAATCGGTCAAGGGGTCTGATGTGTGGTAGCTGCAGGTGGAGTCCAGGTGTGTCTGTAGCT  
ACTTTGGCTCCTCGGCCCCCGGGCCCCGTTTCCCAAAACAGAAGCTTCCAGGCGCTCTCTGGGCTTCATCCCGCCATCG  
GGCTTGGCCGAGGTCCACAGTCTGATCGGAAGAAACAAGTCCAGCTCTGGCCGGGGCAGGCCACATTTGTGGGCTC  
25 ATGCCCTCTCCTCTGCCGGCAG

#### Intron 7 (SEQ ID NO 11)

GTCTGGGCACTGCCCTGCAGGGTTGGGCACGGACTCCAGCAGTGGGTCTCCCTGGGCAATCACTGGGCTCATGACCG  
GACAGACTGTGGCCCTGGGGGGCAGTGGGGGAATGAGCTGTGATGGGGGCATGATGAGCTGTGTGCCCTTGGCGAAATC  
30 TGAGCTGGGCCATGCCAGGCTGCGACAGCTGCTGCATTAGGCACCTGCTCACGTTTGAAGTGCAGGCTCTCTCCAGTT  
CCGCAAGTGCCTTTGTTTATGATTGCTAAATGCTCTCTGCCAGTTTGTATCTTGAAGCCAAAGGAAAGGTGCCCCCT  
CCTTTAGGAGGGCAGGCCATGTTTGAAGCGTGTCTGCCCCAGCTGGCCCTCAGTGTGGGTCTGAGGCCAAAGGAAACG  
TGTCCCCCTTCTTAGGAGGACGGGCGTGTGAGCCACGCCCCGCTGAGCGGGCTCTCAGTGTGGGTCTGTCCACGT  
GGCCCTGTGGCCCTTTCAGATGTGGTCTGTCCACGTGGCCCTGTGGCTCTTTCAGATGCCTGTAGCACTTGTCCGGC  
35 TCTAGGGGACAGTCTGTGTCCACCATGAGGCTCAGAGACCTCTGGGCGAATTCCTTGGCTCCAGGGTGGGGGTGGAG  
GTGGCTGGGCTGTGGGACCCAGACCTGTGCCCGGAGCTGGGCAGCAACTCCTGGATCACATATGCCATCCGGGCCA  
CGGTGGGCTGTGTGGGTGTGAGCCAGCTGGACCCACAGGTGGCCAGAGGAGACGTTCTGTGTACACACTCTGCCTAA  
GCCCATGTGTGTCTGACAGAGACTCGGCCCGGCCAGCCACAGTGGCCCTGCATTCCAGCCCGAGCCCGCACTTCATCACA  
AAGTGTGACCCAAAGGGACGGAGGGTCTTGGCCACGTGGTCTGCTGTCTCAGCACCACCGGCTCACTCCCATGTG  
40 TCTCCCGTCTGCTTTCGAG

TGTAGTACAGGTGGCCAGGTGCCATTGCGCTCGGGTGCGTGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCTCTCTGCGC  
 CTTCCCCCACTGNCCTTCTGCCCCGGGCCACAGAGTCTCCTTTTCTGCCCCCGCCCCCTCCGGCTCCTGGGCTGCAGGC  
 TCCCAGGCGCCCGGAAACATGGCTCGGCTTGCGGCAGCCGGAGCGGAGCAGGTGCCACACGAGGCCTGGAAATGGCAAGC  
 GGGGTGTGGAGTTGCTCCTGCGTGGAGGACGAGGGGCGGGGGTGTGTCTGGGTGAGTGTGCGCCGAGCGTTTGAGCCT  
 GCAGCTTGTGAGTCCAAGTTACTACTGACGTGGACACCCGGCTCTCACACGCTTGATATCTCTCTCTCCCGATACAAA  
 GGATTTTATCCGATTCTCATTCCTGTCCCTGTGCTGTGACCCCCGCGAGGGGCGGGGCTCTTCTCTCTGTGACTAGATT  
 CCCATCTGGAAGTGCAGGGTTGACCGGTGTAGTTTGCTCCTCTCGGGGGGCTGTGGTGCCATGGGGCAGGCGGCCTGG  
 GAGAGCTGCCGTACACAGCCACTGGGTGAGCCACACTCACGGTGGTAGAGCCACAGTGCCTGGTGCCACATCACGTCTT  
 CTGGATTTTAAGTAAACACACACCTCCCGGAGGCATCTGCTGCGACCCTGTGTGTGCTGGGGAGAGTGGTAGCAC  
 GGAGGAAATTCTGTGCACACTCAAGGTCTCAGCAAGGTCTCCGAGTCAGGTGGAACGTGGAGGCCTCTCTCTGGGATC  
 GTCTCCAGCGGATAAAGGACTGTGCACAGCTTCGGAAGCTTTTATTAAAAATATAACTATTAATTATTGCATTATAAGT  
 AATCACTAATGGTATCAGCAATTATAATATTTATTAAAGTATAATTAGAAATATTAAGTAGTACACAGTTCTGGA AAAA  
 CACAAATTGCACATGGCAGCAGAGTGAATTTTGGCCGAGGGACAGTGTGCACATGTGTGTAAGCGGCCCCAGGCCAC  
 AGAATTGCTGACAAAGTCACCTCCCCAGAGAAGCCACCACGGGCCCTCCTTCGTGGTCGTGAATTTTATTAAGATGGATC  
 AAGTCACGTACCGTCCACGTGTGGCAGGGCTTTGGGGAATGTGAGGTGATGACTGCGTCCTCATGCCCTGACAGACAGGA  
 GGTGACTGTGTCTGTCTGTCCCTAGGACACGGACAGGCCCAAGCTCTAGTCCCCATCGTGGTCCAGTTTGGCCTCTGA  
 ATAAAAACGTCTTCAAACCTGTTGCCCCAAAAAATAAGAACAGAGAGAGTTCCCATCCCATGTGCTCACAGGGGCGTA  
 TCTGCTTGCGTTGACTCGCTGGGCTGGCCGGACTCCTAGAGTTGGTGCGTGTGCTTCTGTGCAAAAAGTGCAGTCTCTT  
 GCCCATCACTGTGATATCTGCACCAGCAAGSAAAGCCTCTTTTCTTTTCTTTCTTTTCTTTTGGAGACGGAACGTCA  
 CTGTTGTCTGCTGGGCTTGAGTGCAGTGGCGCATCTCAACTCACTGCAACCTCCGCCTCCGGGTTCCAGCATTTCTC  
 CTGCTCAGCCTCCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCGCTGGCTAATTTTGTATTTTGTAGTAGAG  
 GGGTTTTTGCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGTG  
 GGATTACAGGTGTGAGCCATCACGCCAGCCGGAAGCCTCTTTTAAAGGTGACCACCTATAGCGCTTCCGAAAATAAC  
 AGGTCTGTGTTTTTGAGTAGGCTGCAAGCGTCTCTTAGCAACAGGAGTGGCGTCTGTGGGCTCTGGGGATGGCTGAGGG  
 TCGCGTGGCAGCCATGCCTTCTGTGTGCACCTTTAGGTTCCACGGGGCTATTCTGCTCTCACTGTTTGTCTGAAAACGCA  
 CCTTGGCATCCTTGTGTTGGAGAGTTTCTGCTTCTCGTTGGTCTGCTGAAACTAGGGGCAAGGTTGTATCCGTTGGCGC  
 GCAGCGGCTACATGTAGGCTCATGAGTCTTACCCTGGACAAAATCTTGAAAAA AAAAAAGGAGTCCGGTTAAGCAT  
 TCATTCCGGGTCAAGTGTCTGGTTCTGTGAATAAACTCTAAGATTAAAGAAACCTTAATGAAGAAAACCTTGATGATT  
 AGAGCAAGGATGTGGTCACACCTGTGGCTGGATCTGTTTACGCGCCCCAGTGTCAGGTGAGAGTGGGGAGCAGGGATTG  
 TTTGTTACAGAGTCTCATCTGGTATGTTTCTGAGGTGTTTGGCGGCTGAATGGTAGACGTGTGTTTGTGTGTATGAGGT  
 TCTGTGTCTGTGTGGCTCGGTTTGAGTGTACGCATGTCCAGCACATGCCCTGCCCGTCTCTCACCTGTGTCTTCCCGC  
 CCCAG

GTGAGGCCCTCCTCTTCCCCAGGGGGGCTTGGGTGGGGGTGATTGTCTTTTGATGCATTAGTGTTAATATTCTGGTGCTCTGGAGACCATGACTGCTCTGTCTTGAGGAACGACAGAGGTGTCAGCCCCCTTCTTGGTATGAAGCCGCACGGGAGGGGTTGCACAGCCTGAGGACTCGCGGCTCCACGCAAGGCTCTGTCCAGCGGCCATGTCCAGAGGCCCTCAGGCTCAGCAGGCGGGGGCCGCTGCCCTGCATGATGAGCATGTGAATTCAACACCGAGGAAGCACACCAGCTTCTGTACCGTCACCCAGGTTCT

CGTTAGGGTCCTTGGGGAGATGGGGCTGGTGCAGCCTGAGGCCCCACATCTCCAGCAGGCCCTCGACAGGTGGCCTGGA  
CTGGGGCGCCTCTTCAGCCCATGCCCATCCCACTTGCATGGGGCTACACCCAAGGACGCACACCTAAATATCGTGCC  
AACCTAATGTGGTTCAACTCAGCTGGCTTTTATTGACAGCAGTTACTTTTTTTTTTTAATACTTTAAGTTCTAGGGTAC  
ATGTGCACGACGTGCAGGTTAGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACTCATCATTTACA  
5 TTAGGTATATCTCCTAATGCTATCCCTCCCCACTCCCCCATCCCATGACAGGCCCTGGTGTGTGATTTCCCCACCCTG  
TGTCGAAGTGTTCTCATTGTTCAAGTCCCACCTGTGAGTGAGAACATGTGGTGTGTTGGTTTCTTTCCTTGCAATAGTTT  
GCTCAGAGTGATGGTTTCCAGCTTCGTCCATGTCCCTACAAAGGACATGAACTCATCCTTTTTTATGACTGCATAGTATT  
CCGTGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATCGATGGACATTTGGGTTGGTTGCAAGTCTTTGCTACT  
GTGAATAGTGCCGCAATAAACATACGTGTGCATGTGCTTTATAGCAGCATGATTATAATCCTTTGGGTATATACCCAG  
10 TAATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATCACCACACTGTCTTCACAATGGTTGAA  
CTAGTTTACACTCCCAACAGTGTAAAAGTGTCTGGTGTGAGAGGATGTGGACAGCAGTTATTTTTTATGAAAA  
TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTTCTCTCGAAGAC  
TCCGGGTTTTTCTGTGCATCTTTTGAACCTCTAGCTCCAATTATAGCATGTACAGTGGATCAAGGTTCTTCTTCATTAA  
GGTTCAAGTTCTAGATTGAAATAAGTTTATGTAAACAGAAACAAAAATTTCTGTACACACAACCTGTCTGGGATTGGA  
15 GGAAAGTGTCCTCGAGCTGGCGGCACACTGGTCAGCCCTCTGGGACAGGATACCTCTGGCCCATGGTCATGGGGCGCTGG  
GCTTGGGCCTGAGGGTCACACAGTGCACCATGCCCAGCTTCTGTGGATAGGATCTGGGTCTCGGATCATGCTGAGGACC  
ACAGCTGCCATGCTGGTAAAGGGCACCAGTGGCTCAGAGGGGGCGAGGTTCCAGCCCCAGCTTTCTTACCGTCTTCAG  
TTATTTTTCCCTAAGAGTCTGAGAAGTGGGGCCGCGCTGATGGCCTTGGTTCGTCTTCAGCTGGCACAGAATTGCACAA  
GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACTGTAGAGAGCTCGTCTGTGGAAA  
20 GAAATTTAAGTTTTTCATTTAACCGCTTTGGAGAATGTTACTTTATTTATGGCTGTGTAATGTTTGACATTCACTCCC  
TCGTAGACAGATACTACGTAAAAAGTGTAAGTTAACCTTGCTGTGATTTTTCCCTTATTTTAG

## Intron 10 (SEQ ID NO 14)

GTGAGGCCCGTGCCGTGTCTGTGGGACCTCCACAGCCTGTGGGCTTGCAGTTGAGCCCCCGTGTCTGCCCCCTGG  
25 CACCGCAGCGTTGTCTCTGCCAAGTCTCTCTCTGCGGTGCTGGATCCGCAAGAGCAGAGGCGCTTGGCCGTGCACC  
CAGGCCTGGGGGCGCAGGGGCACCTTCGGGAGGGAGTGGGTACCGTGCAGGCCCTGGTCTCTGAGAGACGCACCCAGGTT  
ACACACGTGGTGAGTGCAGGCGGTGACCTGGCTCCTGTGCTCTTTGGAAAGTCAAGAGTGGCGGCTCTGGGGCCCCAG  
TGAGACCCCCAGGAGCTGTGCACAGGGCTGCAGGGCCGAGGCGGAGCCTCCTCCCCAGGGTGACCTGAGCCTGCGGA  
GAGCAGGAGCTGCTGAGTGAGCTGGCCACAGCGTTGCTGCGGTACGTTCTGCGTGGGGTTGTTTGGGATCGGTGGG  
30 AGAATTTGGATTTGCTGAGTGCTGCTGCTTGAACACGGAGATGGCTAGGAGTGGGTTTTCAGAGTTGATTTTGTGAAT  
CAAACTAAAAATCAGGCACAGGGGACCTGGCCTCAGCACAGGGGATTGTCCAATGTGGTCCCCCTCAAGGGCGCCCCACAG  
AGCCGGTGGGCTTGTTTTAAAGTGCATTTGACGAGGGACGAGAAACCTTGAAAGCTGTAAAGGGAACCTTCAGAAAAATG  
TGGCCGCAGGGGTGGTTTCAGGTGCTTTGCTGGGCTGTGTTTGTGAAAACCCATTGAGACCGCCCTCCAAGTCCACCC  
TCCAGGTCCACCTCCAGGGCCGCCCTGGGCTGGGGGTATGCTGCGGTTCTTGTGCGCAGCCCCGAGCACAGCAGGC  
35 TGTGCACATTTAAATCCACTAAGATTCACTCGGGGGAGCCAGGTCCCAAGCAACTGAGGGCTCAGGAGTCTGAGGCT  
GCTGAGGGGACAGAGCAGACGGGGAACGTGCTTCTGTGTGGCAAGTTCTGAGGGTGTGGCCAGGAGGTGGCTCAGA  
GTGTATGTTGGGTCCTCCCGGGGGCAGAACTCTGTCTGTATGAGTGGCAGCCATGTAAAGGAAGGGGTGGCCACAG  
GGAGCTGGGAATGCACAGGGGAGCTGCGCAGCTGGCCGAGGTCCAGGGCCAGGCCACAGGAAGGGCAGGGGGACGCCC  
GGGGCCACAGCAGAGGCCGAGGAAGGGAAGGGATGCCAGGCCAGAGCAGAGGCTACCGGGCACAGGGGGCTCCCTG  
40 AGCTGGGTGAGCAGGCTCATGACTCGGCAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTGCCAGCTCACAG

CCCAGCCAGGTCCTCCGCGCTGAGCAGGAACCTCAGAACCTCCCTTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT  
AGGAGAAAAACAGGCAAAGTCGTTGAGAAAACGTCCTAAAAGAAGGTGGGATGGTGGCAATTTCTGTCCAGATTTTAGTCT  
GCCCCGGACCACAGATGAGTCTATAACGGGATTGTGGTGTGGCATGGGGACACATGAGATGGACCATCACAGAGGCCAC  
TGGGGCTGCACCTCCCATCTGAGTCCTGGCTGTCCCGGGTCCAGGCCAGGTTCTTGATGCTCACCTACCTGTCTGTCCC  
5 GGGAGACAGGGAAAGCACCCCGAAGTCTGGAGCAGGGCTGGGTCCAGGCTCCTCAGAGCTCCTGCCAGGCCAGCACCTT  
GCTCCAAATCACCACCTTCTCTGGGGTTTTCCAAAGCATTTAACAAGGGTGTGAGGTTACCTCCTGGGTGACGGCCCCGCA  
TCCTGGGGCTGACATTGCCCTCTGCCTTAG

**Intron 11 (SEQ ID NO 15)**

10 GTGAGCGCACCTGGCCGGAAGTGGAGCCTGTGCCCGGCTGGGGCAGGTGTGCTGCAGGGCCGTTGCGTCCACCTCTGCT  
TCCGTGTGGGGCAGGCGACTGCCAATCCCAAAGGGTCAGAGGCCACAGGGTGCCCTCGTCCCATCTGGGGCTGAGCAGA  
AATGCATCTTTCTGTGGAGTGAGGGTGCTCACAACGGGAGCAGTTTTCTGTGCTATTTTGGTAAAAGGAAATGGTGAC  
CAGACCTGGGTGCACTGAGGTGTCTTCAGAAAGCAGTCTGGATCCGAACCCAAGACGCCCGGGCCCTGTCTGGGCGTGAGT  
15 CTCTCAAACCCGAACACAGGGGCCCTGTGGGCATGAGTCCCTCTGAACCCGAGACCTGGGGCCCTGTCTGGGCGTGAGT  
CTCTCCGAACCCAGAGACTTCAGGGCCCTTTTGGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC  
AGTCTACAGGATGCCATGAGTTCATGATCACGTGTGACCCATCAGGGGACAGGGCCATGGTGTGGGGGGGTCTCTACAA  
AATTCTGGGGTCTGTTCCTCCAGAGCCGAGAGCTCAAGGCCCGTCTCAGGCTCAGACACAAATGAATTGAAGATGGA  
CACAGATGCAGAAATCTGTGCTGTTTCTTTATGAATAAAAAGTATCAACATTCCAGGCAGGGCAAGGTGGCTCACACCT  
20 ATAATCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACTTGAGGCCAGGAGTTTGAGGCCAACCTAACCAACATAGTG  
AAATTCCATTTCTACTTAAAAAATACAAAAATTAGCCTGGCCTGGTGACACGCCTGTAGTCCCCGCTATGCGGGAGGC  
TGAGGCAGGAGAATCATTTGAACCCAGGAGGCAGAGGTGTCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCA  
ACAGAGTGAGACTTCATCTTAAAAAAAAAAAAAGTATCAGCATTCCAAAACCATAGTGACAGAGTGTTTTTTTATTC  
TGTCTTCGATAATATTACTGGTGCTGTGCTAGAGGCCGGAACCTGGGGTGCCCTCCTCTGAAAGGCACACCTTCATGG  
GAAGAGAAATAAGTGGTGAATGGTTGTTAAACAGAGGTTTAAACTGGGGTCTGTGCTGTTCTGAGTTAACAGTCCAGATC  
25 TGGACTTTGCCTCTTCCAGAATGCTCCCTGGGGTTTGCTTCATGGGGAGCAGCAGGTGTGGACACCCCTCGTGATGGGG  
GAGCAGCAGGTGCAGACGCCCTCATGATGGGGAGTGGCAGGTGCAGACACCTTGTGATGGTGCCAGCATGTCCCTG  
TTGCAGTCCCTCCCCACAAGGATGCCGGTCTCCTGTGCTCCCCACAGTCCCTGCTTCCCTCTCACAGCCTTACCTGGTC  
CTGGCCTCCACTGGCTTTGTCTGATGATTTCCACATTTCTGGGCTCCAGCACCTCTTCGCTCTCCAGGCACCTCT  
GCAGTGCTGGCCATACCAGTCAGCTGTGAAGTGTCCACTGCTTATTTGCTCCCCATGAAATGTATTTTTTAGGACAGGC  
30 ACCCTTGGTTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAAATGG  
GTTCTCTCTAAACACATTGCAAAGCCACAGAGGCTAGTGAGGATGGGTGGGCATCAGGTATCAGATGTGGGTCCAATG  
CCAGAATATTCTGTGCTCCCAAAGGCCACTTGGTCAGAGTGTGTGCTTGAGAGGTGGCTCTAAAAGCTCAGCAGTGGAG  
GCAGTGGTTCGCCATACTCAGGGTGAAGTACATCCTCTGTGCTGAAGTATACAGCAGAGGCTTGAAGGGCATCTGGGA  
GAAGAAAACAGGCAAATGATTAAGAAAAGTGAAGAAAGGAAAGTGGTAAGATGGGAATTTCTGTCCAGATTTTAGTC  
35 TCCCAAACACAGCTCAGATGGTAGAATGTGGTCAGAACTGATGGACAGAACATAGAACAAAACGGAAGCCCTATCTCT  
CAGAAACGTGTGTTAATGTGGTATGTGGCACAGCTGATGAAAAGAGAGTGTGTGTGAATTTTTTTCTGAGAAAACT  
GACTGGAAGCAAATAAGTTGTGCTTTACAGCATATACCAGAGCAGATTCTAGGTAGAAGAGGAGACATGCAACAAC  
ACCAGCAACAGAAATAAAACAAAAGACTCAAAGGGAAGGGAGGTGAACGTTCCCTGGTTTGGTGTGGGGAAGGACACAC  
AGGGAGGCGGATGAAACAGTGAGGCAACGGGCATTGCTTTCAGTGCAGAGAACTCAGCTTGCTGAGCCACAGTGAAA  
40 ATGGCCATTCCCTGGAGCGTTTGTGACGTGATTTATTTAAGGCGCCCTGTGAGGTCTGCACATTCATCCTCTCACTTT



5 GTTCTCCTAACCACTGAGAGGTAGAGGAGAAAGGCTCCAGGGGAGCAGCCGCCCTTGGTCAACCAGCTGGCAAAGGGC  
ATGCATGATTGCAGCCTGGCCTCTGCTCCGGGGCCCTTGCTCTGCCCGAGGACCCACACAAGTCAGACCCATAGGCTC  
AGGGTGAGCCGGAGCCCAAGGTCGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC  
CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC  
10 GAGACCTGTCCCCATCCCTCATGCTGGCTCCTTTCTGGGCTTGCCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAAG  
ACTTTTCTGGAAAGCAGCTGTTTGCATGGAAGTCCTCACAATGTCTGTCTTCCAGTAATTCACCTTCTGAAGTGA  
CCAGACATTATCACGGGTCTTATTTACCATTTCAGTGTTCCAGGCAGGGGGACTTGCCACAGCAAGTCACGAACCTGCC  
CAAATACAGGGCTAAGGAGATATTATGCATCACAAACTTGCTCTGCCATTAAACATTTTCAAAGAATTTTGAAGAAT  
GTTTAATGGCACAAACGTTTATTTCAATGTAGCAGTGTTCAAAGCTGGATGTAAAGAACACACCCAGGAGCCTGCCG  
15 TGAATGTCATGTGTGTTTCTTTGGACATGGACATACATGGGCAGTGAGTGGTGGTGAGGCCCTGGAGGACATCGGTGG  
GATGCCCTCCATCCTGCCCTCTGGAGACACCATGTGTGCCAGTGCACCTCACTGGAGCCCTGTTAGCTGGTGCCACCTG  
GCTCTTCCATCCCTGAGATTCAAACACAGTGAGATTCCCCACGCCCACTCAGTGTTCTCCACAAAAAACCCTGAGTCAC  
ACCTGTGTTCACTCGAGGGACGCCCGGGAGCCAGGGCTCCACAGTTTATTATGTGTTTTTGGCTGAGTTATGTGCAGATC  
TCATCAGGGCAGATGATGAGTGACAAACACGGCCGTGCGAGGTTTGGATACACTCAACATCACTAGCCAGGTCCTGGTG  
20 GAGTTTGGTCATGCAGAGTCTGGATGGCATGTAGCATTGGAGTCCATGGAGTGAGCACCCAGCCCTCGGGCTGCAGC  
GCATGCCCCAGGCAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTCTTTGGAGCAAGCTTTCAGGAGGGGGCTGGGTGT  
GGGGCAGGCACCTGTGTCTGACATTCCCCCTGTGTCTCAG

#### Intron 12 (SEQ ID NO 16)

20 GTGAGCAGGCTGATGGTCAGCACAGAGTTCAGAGTTCAGGAGGTGTGTGCGCAAGTATGTGTGTGTGTGTGCGCGCT  
GCCTGCAAGGCTGATGGTGACTGGCTGCACGTAAGAGTGACATGTACGCATATACAGTGAGCACATACATGTGTGCAT  
GTGTGTACATGAAGCATGGCAGTGTGTGCACAGGTGTGCAAGGGCACAAGTGTGTGCATGCGAATGCACACCTGACA  
TGCAATGTGTGTTCTGTGCACAGTCGTGTGGGCATTACGTGAGGTGCATGCGTGTGGGTGTGAGTGTGAGTAGCATGTGT  
GCACATAACATGTATTGAGGGGTCTCGTGTTCACCCCGCTAGGTCTCAGCACCAAGTGCCACTCCTTACAGGATGAGAC  
25 GGGGTCCCAGGCCTTGGTGGGCTGAGGCTCTGAAGCTGCAGCCCTGAGGGCATTGTCCCATCTGGGCATCCGCGTCCACT  
CCCTCTCCTGTGGGCTTCTGTGTCCACTCCCCCTCTCCTGTGGGCATTTACATCCACTCCACTCCCTCTCTCCTGTGGGC  
ATCCGCGTCCACTCCCCCTCTGTGTGGGCATCTGCGTCCACTCCCCCTCTGTGTGGGCATTTGCGTCCACTCCCTCTCCT  
GGTTCCTTCTGTCTTGGCCGAGCCTCGGGGGCAGGCAGATGACACAGAGTCTTGACTCGCCAGGGTGGTTGCGCAGCTG  
CCGGGTGAGGGCCAGGCCGATTCTACTGGGAAGAGGGATAGTTTCTGTCAAATGTTCTCTTTCTTGTTCATCTGA  
30 ATGGATGATAAAGCAAAAAGTAAAAACTTAAATCCCAGAGAGGTTTCTACCGTTTCTCACTCTTCTTGGCGACTCTAG

#### Intron 13 (SEQ ID NO 17)

GTGAGCCGCCACCAAGGGGTGCAGGCCAGCCTCCAGGGACCCTCCGCGCTCTGCTCACCTCTGACCCGGGGCTTCACTT  
TGGAACCTCTGGGTTTTAGGGGCAAGGAATGTCTTACGTTTTTCAAGTGGTGCTGCTGCTGTGCACAGTCTGTTCGCGTG  
35 GCTCTGTGCAAAGCACCTGTTCTCCATCTCTGGGTAGTGGTAGGAGCCGGTGTGGCCCCAGGTGTCCCCACTGTGCTGT  
GCACTGGCCGTGGGACGTCTAGGAGCCATCCCAGGGCAGCAGGGGCATGGGGTAAAGAGATGTTTATGGGGAGTCTTAG  
CAGAGGAGGCTGGGAAGGTGTCTGAACAGTAGATGGGAGATCAGATGCCCGAGGATTTGGGGTCTCAGCAAAGAGGGCC  
GAGGTGGGTGCAGGTGAGGGTCGTGGCCCCACCCCGGAAGGTGCAGCAGAGCTGTGGCTCCCCACAGCCCCGCCA  
GCACCTGTGCTCTGGGCATGGCTGTGCTCCTGGAACGTTCCCTGTCTGGCTGGTCAGGGGGTGGCCCTGCCAAGAATCG  
40 ACAACTTTATCAGAGGGAAGGGCCAATCTGTGGAGGCCACAGGGCCAGCTTCTGCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTCGGGGCTGTACCAAAGGGCAGTCGGGCACCACAGGCCCGGGCCTCCACCTCAACAGGCCTCCCAGGCACTG  
GGAGCTGAATGCCAGGAGGCCGAAGCCCTCGCCCATGAGGGCTGAGAAGGAGTGTGAGCATTGTGTACCCAGGGCCG  
AGGCTGCGCGAATTACCGTGACACTTGATGTGAAATGAGGTCGTCGTCTATCGTGAAACCCAGCAAGGGCTCACGGGA  
GAGTTTTCCATTACAAGGTCGTACCATGAAAATGGTTTTTAACCGAGTGCTTGCGCCTTCATGCTCTGGCAGGGAGGGC  
5 AGAGCCACAGCTGCATGTTACCGCCTTTGCACCAGCTCCAGAGGCTTGGGACCAGGCTGTCTCAGTTCCAGGGTGCCTCC  
GGCTCAGACCGCCCTCCTCTCTGCCTTCTCTCTGCTCAAATCTTCCCTCGTTTGATCTCCCTGACGCGTGCTGGG  
CCCTCGTGCAAGCTGCTTGACTCCTTTCCGGAACCCCTTGGGGTGTGCTGGATACAGGTGCCACTGAGGACTGGAGGTGT  
CTGACACTGTGGTTGACCCAGGTCAGCTGGCGTGCTTGGGGCCTCCTTGGGCCATGATGAGGTGAGGAGTGTTCCTC  
CAGGTGAAAACCTCTGGGAACTCCAGGGCCATGTGACCTGCCACCTGCTCCTCCCATATTAGCTCAGTCTTGTCTCTC  
10 ATTTCCCAACCAGGGTCTCTAGCTCCGAGGAGCTCCCGTAGAGGGCTGGGCTCAGGGCAGGGCGGCTGAGTTTCCCCAC  
CCATGTGGGGACCTTGGGTAGTCGCTTGATTGGGTAGCCCTGAGGAGGCCGAGATGCGATGGGCCACGGGCCGTTTCCA  
AACACAGAGTCAGGCACGTGGAAGGCCAGGAATCCCTTCCCTCGAGGCAGGAGTGGGAGAACGGAGAGCTGGGCCCCG  
ATTTACGGCAGCCAGGCTGCAGTGGGCGAGGCTGTGGTGGTCCACGTGGCGCTGGGGCGGGGTCTGATTCAAATCCGC  
TGGGGCTCGGCCTTCTGGCCCGTGCTGGCCGCGCTCCACACGGGCTTGGGGTGACGCCCGACCTCTAGCAGGTGGC  
15 TATTTCTCCCTTTGGAAGAGAGCCCTCACCATGCTAGGTGTTTCCCTCCTGGGTGAGGAGCTGGCCGTGTGGCAACC  
CCGGGACCTTAGGCTTATTTATTTGTTTAAACATTCTGGGCCTGGCTTCCGTTGTTGCTAAATGGGGAAAAGACATCC  
CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGGTCTGGCTTGACTGGTGTGATCTCAGGTCAATCCAGAAGTGGCT  
CAGGAAGTCAGTGAGACCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGTGGGT  
GAGGCCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGAGGTCAGACCAGGTAC  
20 ACGGGGCTCTGATCACACGCACATATGAGCACATGTGCACATGTGCTGTTTCATGGTAGCCAGGTCTGTGCACACCTGC  
CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGGAGGTGACAGGGCTGGCGCGGTGGCTCACACCTGTAGTCCCAGCA  
CTTTGGGAGGCCGAGGCGAGAGGATCCCTTGAGCCAGGAGTTTAAGACCAGCCTGAGCAACATAGTAGAACCCCATCTC  
TATGAAAAATAAAACAAAAATTAGCTGAACATGGTGGTGTGCGCCTGTAGTTCCAATACTTGGGAGGCTGAAGTGGGAG  
GATCACTTGAGCCAGGAGGTGGAAGCTGCAGTGAGCTGAGATTGCACCACTGTACTGCAGCCTGGGTGACAGAGTGAGA  
25 GCCCATCTCAACAACAACAAGAAGACTGACAAATGCAGTTTCTTGGAAGAAACATTTAGTAGGAACTTAACCTACACA  
CAGAAGCCAAGTCGGTGTCTCGGTGTCAGTGAGATGAGATGATGGGTCTCACACCATCACCCAGACCCAGGGTTTATG  
CACCAAGGGGCGGGTGGCTCAGAAGGGATGCGCAGGACGTTGATATACGATGACATCAAGTTGTCTGACGAAGGGCAG  
GATTTCATGATAAGTACCTGCTGGTACACAAGGAACAATGGATAAACTGGAACCTTAGAGGCCTTCCCGAACAGGGGCT  
AATCAGAAGCCAGCATGGGGGCTGGCATCCAGGATGGAGCTGCTTCAGCCTCCACATGCGTGTTCATACAGATGGTGCA  
30 CAGAAACGAGTGATACCTGTGCACACACAGACACGAGCTACTCGCACACACAAGCACACACAGACATGCATGCATGC  
ATCCGTGTGTGTGCACCTGTGCCCATGAGGAACCCATGCATGTGCATTATGCACGCACACAGGCACCGGTGGGCCCAT  
GCCACACCCACGAGCACCCTCTGATTAGGAGGCCTTTCTCTGACGCTGTCCGCCATCCTCTCAG

**Intron 14 (SEQ ID NO 18)**

35 GTATGTGCAGGTGCCTGGCCTCAGTGGCAGCAGTGCCTGCCTGCTGGTGTAGTGTGTGAGGAGACTGAGTGAATCTGGG  
CTTAGGAAGTTCTTACCCCTTTTCGCATCAGGAAGTGGTTTAACCCAACCACTGTGAGGCTCGTCTGCCCCCCTCTCGT  
GGGGTGAGCAGAGCACCTGATGGAAGGACAGGAGCTGTCTGGGAGCTGCCATCCTTCCACCTTGCTCTGCTGGGGAA  
GCGCTGGGGGCTGGTCTCTCTGTTTGGCCATGGTGGGATTGGGGGGCTGGCCTCTCCTGTTTGGCCCTGTGGTGG  
GATTGGGCTGTCTCCCGTCCATGGCACTTAGGGCCCTTGTGCAAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCAG  
40 GCTACCCACCCCTCTCAGGAGCAGAGGCCGCTATCACCACGACAGAGCCCGCGCGCTCTCTGCTTCCAGTACCCG

TCCTCTGCCCCCTGGACACTTTGTCCAGCATCAGGGAGGTTTCTGATCCGTCTGAAATTCAAGCCATGTCGAACCTGCGGT  
CCTGAGCTTAACAGCTTCTACTTTCTGTCTTTCTGTGTTGTGGAAATTTACCTGGAGAAGCCGAAGAAAACATTTCTG  
TCGTGACTCCTGCGGTGCTTGGGTGCGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGTGGGCAGCTTTCCG  
GTGTCTCCTGGGAGGGGAGCTGGGCTGGGCCTGTGACTCCTCAGCCTCTGTTTCCCCCAG

5

#### Intron 15 (WEQ ID NO 19)

GCAAGTGTGGGTGGAGGCCAGTGCAGGGCCCCACCTGCCAGGGGTATCCTTGAACGCCCTGTGTGGGGCGAGCAGCCTC  
AGATGTGTGCTGAAGTGCAGACGCCCGGGCCTGACCTGGGGCCTGGAGCCACGCTGGCAGCCCTATGTGATTAAACG  
CTGGTGTCCCCAGGCCACGGAGCCTGGCAGGGTCCCCAATTCTTGAACCCCTGCTTCCCATCTCAGGGGCGATGGCTCC  
CCACGCTTGGGAGCCTTCTGACCCCTGACCTGTGTCTCTCACAGCCTCTTCCCTGGCTGCTGCCCTGAGCTCCTGGGGT  
CCTGAGCAAGTTCTCTCCCGCCCCCGCCTCCAGCGTCACTGGGCTGCCTGTCTGCTCGCCCGGTGGAGGGGTGTCTG  
TCCCTTCACTGAGGTTCCACACAGCCAGGGCCACGAGGTGCAGGCCCTGCCTGCCCGGCCACCCACACGTCCTAGGAGGG  
TTGGAGGATGCCACCTCTGGCCTCTTCTGGAACGGAGTCTGATTTTGGCCCCGAG

10

#### 3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATGTTTGAATCCTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGGTGGT  
GGTCAGTGCAGGGCCCATGGCCTGGCTGTGCATTACGGAAGTCTATGAGTGAATGGGGTGTGGTGTGAGTGCAGGGCCCATG  
GCCTGGCTGGGCCTGGGAGGTTTCTGATGTGTGAGGCAGGAGGGGAAGGAGGGTAGGGGATAGACAGTGGGAGCCCCCA  
CCCTGGAAGACATAACAGTAAGTCCAGGCCGAAGGGCAGCAGGGATGTGGGGGCCAGCTTGGGCGCGGGGATGATG  
GAGGGCCTGGCCAGGGTGGCAGGGATGATGGGGCCCCAGCTGGGGTGGCAGGGGTGATGGGGGGGCTGGTCTGGGTGG  
CGGGGAAGATGGGAAGCCTGGCTGGGCCCCCTCCTCCCTGCCTCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT  
GGTGACATCCTCTGGGCCATCAGCTTTCATGGAGGTGGGGGGCAGGGGCATGACACCATCCTGTATAAAATCCAGGATT  
CCTCCTCCTGAACGCCCACTCAGGTTGAAAGTACATTCCGCCTCTGGCCATTCTCTAAGAGTAGACCAGGATTCTG  
ATCTCTGAAGGGTGGGTAGGGTGGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGGTGGGGCTGGTATGCTCTCTC  
ATCCTCTTATCATCTCCAGTCTCATCTCTCATCTTATCATCTCCAGTCTCATCTGCTTCTCTTATCTCCAGT  
CTCATCTGTCTCTTACCATCTCCAGTCTCATCTTATCCTTATCTCCTAGTCTCATCCAGACTTACCTCCCA  
GGGCGGGTGCCAGGCTCGCAGTGGAGCTGGACATACGTCTTCTCAGGCAGAGGAAGTGAAGGATTGCAGAGAACAG  
GAGGGCGGGCTCAGAGGGACGCAGTCTTGGGTGAAGAAACAGCCCTCCTCAGAAGTTGGCTTGGGCCACACGAACCG  
AGGGCCCTGCGTGAGTGGCTCCAGAGCCTTCCAGCAGGTCCCTGGTGGGGCCTTATGGTATGGCCGGGTCTACTGAGTG  
CACCTTGGACAGGGCTTCTGGTTTGTGAGTGCAGCCCGACGTGCCTGGTGTGGGGTGGGGGCTTATGGCCACTGGATATG  
GCGTCATTTATTGCTGCTGCTTCAGAGAATGTCTGAGTGACCGAGCCTAATGTGTATGGTGGGCCAAGTCCACAGACTG  
TGTCTGTAATGCACTCTGGTGCCTGGAGCCCCGTATAGGAGCTGTGAGGAAGGAGGGGCTCTTGGCAGCCGGCCTGGGG  
GCGCCTTTGCCCTGCAAACTGGAAGGGAGCGGCCCGGGCGCGCTGGGCGGACGACCTCAAGTGAGAGGTTGGACAGAAC  
AGGGCGGGGACTTCCAGGAGCAGAGGCCGCTGCTCAGGCACACCTGGGTTTGAATCACAGACCAACaGGTCAGGCCATT  
GTTACAGCTATCCATCTTCTACAAAGCTCCAGATTCTGTGTTTCTCCGGGTGTTTTTGTGTAATTTTACTCAGGATTACT  
TATATTTTTTGTCTAAAGTATTAGACCCTTAAAAAAGGTATTTGCTTTGATATGGCTTAACCTACTAAGCACCTACTTTAT  
TTGTCTGTTTTTATTATTATTATTATTATTATTAGAGATGGTGTCTACTCTGTCAACCCAGGTTGTTAGTGACGTGGCAC  
AGTCATGGCTCGCTGTAGCCGCAAAACCCAGGCTCAAGTGATCCTCCGGCCTCAGCTTCCAGAGTGTCTGGGATTACAG  
GTGTGAGCCACTGCCCTTGCCTGGCACTTTAAAAAACCACTATGTAAGGTCAAGTCCAGTGGCTTCCACACCTGTCTATCC  
CAGTAGTTTGGGAAGCCGAGGCAGAAGGATTGTCTGAGGCCAGGAGTTTGAGACCAGCATGGGTAACATAGGGAGACCCC

20

25

30

35

40

ATCTCTACAAAAATGCAAAAAGTTATCCGGGCGTGGGGTCCAGCATCTGTAGTCCCAGCTGCTCGGGAGGCTGAGTGGG  
AGGATCGCTTGAGCCCGGAGGTCTGGCTGCAGTGAGCTGTGATTGTACCATCGCACTCCAGCCTGGGCAACAGAGTGA  
GACCTGTCTCAAAAAAAAAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGAAGAAGGAAGAAGGAAAGAGAAGAAGAAG  
5 GAAGAAGGAAGAAGAAGGAGAAGGAGGCGCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAAATGAAAATAACA  
AAGTTTTAAAGGGAAGAAAAACCCAGCTCTTTGGACTTCCTTAGGCCTGAACTTCATCTCAAGCAGCTTCCTTCCACA  
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA  
GCCAGGACCCCTGAAAGGGAGTGGTTGTTTTCTGCCTCAGCCCCAGCTCCTGCCGGTCTGCACTGCTGTAACCGTC  
GATGTTGGTGCCAGGTGCCACCTGGGAAGGATGCTGTGCAGGGGGCTTGCCAACTTTGGTGGGTTTCAGAAGCCCCAG  
GCACCTGTGGCAGGCACAATTACAGCCCTCCCCAAAGATGCCACGTCCTTCTCCTGGAACCTGTGAATGTGTACCCG  
10 CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTCATCCTGGATTATCTGG  
TGGGCTGATATGGCCACAAGGGTCCCTAGAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC  
CACTGGCCACTGCTGGCTTTGAGATGGAGAGGGGGTCCCCAGCCAAGGAATGGGGGCAGCCGCTCCATGCTGGAAAAGC  
AAGCAATCCTCCCCGGTCCCTGAGGGCACAGGCCCTGCCACGCCTCGATTTAGGCCAGTGGGACCTGTTTCAGCTTTC  
CGGCCTCCAGAGCTGTAAGATGATGCGTTTGTGTTAGCCACTAAGCTGCAGTGATTGTCACAGCAGCAAAATGGAATAG  
15 CAGTACAGGGAATGAATACAGGGACAGTTCTCAGAGTGACTCTCAGCCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

15

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorengeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

20

25

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalisierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

30

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, läßt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

### Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA<sup>+</sup>-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur  
10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv  
markierter Primer (5'-GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10<sup>5</sup>  
cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl  
inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL),  
2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5 µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl  
15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser  
erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die  
Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach  
Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden  
2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150 µl  
20 Phenol/Chloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-  
Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min  
bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol  
gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stopplösung gelöst. Nach 5  
min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel  
25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom  
ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in  
Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA<sub>n</sub>TTGT)  
30 repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA<sub>n</sub>Na/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

5 In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-loosen Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei  
10 anderen TATA-loosen Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

#### Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

20 Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem  
25 Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer  
30 zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer



angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-CTAGGCCGATTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCAGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprinten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

### Beispiel 8

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

### Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTTCC-3', Position -4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaares PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGGATTTGCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die  
5 verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und  
10 anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 $\alpha$ -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für  
15 die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

#### Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection  
25 bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

30

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

- 5 CHO-K1 und HEK 293 Zellen wurden in wassergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO<sub>2</sub> kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit  
10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszu-plattieren.

- Für die Transfektion wurden pro Well jeweils  $2 \times 10^5$  –HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium  
15 entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serum-freies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium  
20 hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H<sub>3</sub>PO<sub>4</sub>; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl  
25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO<sub>3</sub>)<sub>4</sub> Mg(OH)<sub>2</sub>·5H<sub>2</sub>O; 2,67 mM MgSO<sub>4</sub>; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

- 30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysats und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl<sub>2</sub>;

50 mM  $\beta$ -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100  $\mu$ l 1 M  $\text{Na}_2\text{CO}_3$  gestoppt und die Absorption bei 420 nm bestimmt.

- 5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK  
10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im  
15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

## Literaturverzeichnis

- 5 Allsopp, R. C., Vazire, H., Pattersson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.
- 10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Wiley-Intersciences, New York.
- Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.
- 15 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.
- Counter, C. M., Ailion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.
- 20 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Ailion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.
- 25 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. Mol. Cell. Biol. 14:4894.
- 30 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.
- Harley, C.B., Futcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.
- 35

- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.
- 5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.
- 10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.
- Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.
- 15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.
- Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.
- 20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.
- 25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)<sub>n</sub> among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.
- Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.
- 30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.
- 35 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* 57:103-113.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.

5 Shay, J. W. (1997). Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.

10

Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.

15

Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.



**Patentansprüche**

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit.
- 5 2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser Sequenzen handelt.
- 10 3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.
- 15 4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der Ansprüche 1 bis 3.
5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide oder Proteine kodieren.
- 20 6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.
- 25 7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.
8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß einem der Ansprüche 4 bis 6.

9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:
- 5           A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reportergenexpression.
- 10
10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screent.
- 15
11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.
- 20
12. Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:
- A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält
- 25           mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

- C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

Fig. 1

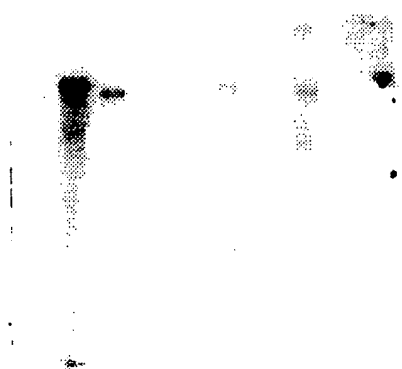
A

1 2 3 4 5 6 7 8 9 10



B

1 2 3 4 5 6 7 8 9 10



**Fig. 2**

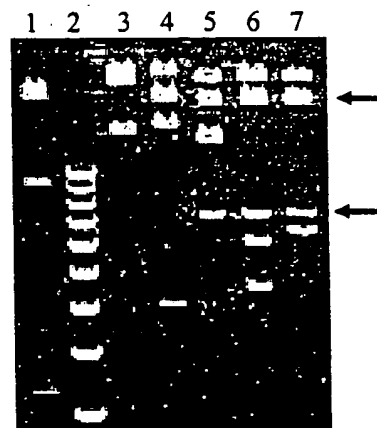


Fig. 3

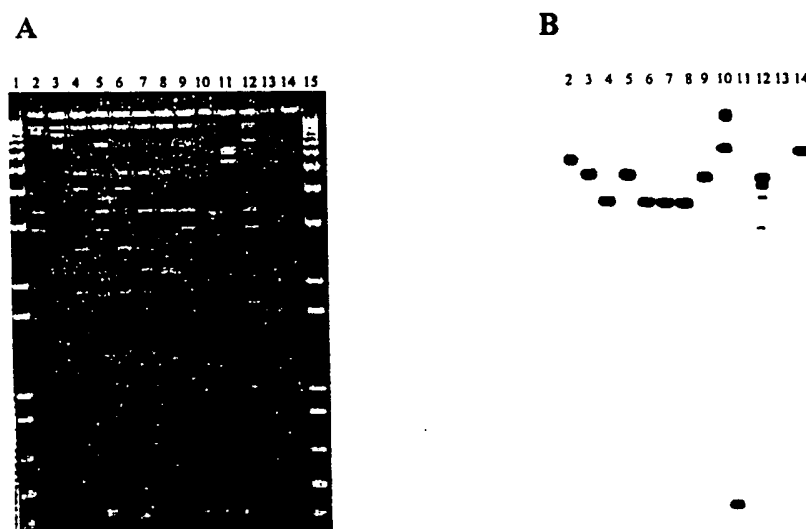


Fig. 4

GAGCTCTGAA CCGTGGAAAC GAACATGACC CTGCTCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA 70  
 GTGGTGTGCA GGAATGGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTTAT 140  
 CATCTTCACC CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAAGCCATGA CAAAACCTCAG 210  
 TACAAACACC ACTCTTTTAC TAGGCCACCA GAGCAGGGGC CACACCCCTG ATATATTAA AGTCCAGGAG 280  
 AGATGAGGCT GCTTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTCC TCTAGACTAG 350  
 TAGACCCTGG CAGGCACTCC CCCAAATTCT AGGGCCTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG 420  
 GAGACTCAGC CTGGGGTGCC ACACTGAGGC CAGCCCTGTC TCCACACCCT CCGCTCCAG GCCTCAGCTT 490  
 TCCAGCAGC TTCCTAAACC CTGGGTGGGC CGTGTTCAG CGTACTGTC TCACCTGTCC CACTGTGTCT 560  
 TGCTCAGCG ACGTAGCTCG CACGGTTCCT CCTCAGATGG GGTGTCTGTC TCCTTCCCA ACACCTACAT 630  
 GCGTGAAGG GAGGAGATTG TCGCCTCC CAGACTGGCT CTCTGAGCCT GAACCTGGCT CGTGGCCCC 700  
 GATGCAGGTT CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCAGGCGC TCCCGCTCTC CTGTCTCTG 770  
 CCGGGGCTG CCGGTGTGTT CTCTGTTC TGCTCTCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCCG 840  
 CTAGGGTCTC GGGGTTTTTA TAGGCATAGG ACGGGGCGT GGTGGGCCAG GCGCTCTTG GGAATGCAA 910  
 CATTTGGGTG TGAAAGTAGG AGTGCCTGTC CTCACCTAGG TCCACGGGCA CAGGCTCTGG GATGGAGCCC 980  
 CCGCCAGGGA CCCGCCCTTC TCTGCCAGC ACTTTCCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT 1050  
 TTCACAAGC ACTAAGCATC CTCTTCCCA AAGACCCAGC ATTGGCACCC CTGGACATT GCCCCACAGC 1120  
 CCTGGGAATT CAGGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCCG CTGTTTTATT 1190  
 TTAATAGCTA CAAAGCAGGG AAATCCCTGC TAAATGTCC TTTAACAAC TGGTTAAAC AACGGGTCCA 1260  
 TCCGACGGT GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG 1330  
 GAATTACGCT GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAG AAAGAATTC 1400  
 ACCCTAGTGC AGGGGAGTGG TTAGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCACGCA 1470  
 CCTTTTACTA AAGCCAGTTT CCTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG 1540  
 TGGGATGGG GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCTGGGCAGG ATAATGCTCT 1610  
 AGAGATGCCC ACGTCTCGAT TCCCCAAAC CTGTGGACAG AACCCGCCCG GCCCCAGGGC CTTTGCAGGT 1680  
 GTGATCTCCG TGAGGACCTT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCCGAAAA GTAATCCAGG 1750  
 GTTCTGGGA AGAGGCGGGC AGGAGGCTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA CTCAGTCTGA 1820  
 GGCTGAAAAG GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCTCCAG AAGCTGGAAA AAGCGGGGAA 1890  
 GGGACCTCC ACGGAGCCTG CAGCAGGAAG CACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC 1960  
 TCCGGCCTCC GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA 2030  
 GCAACAGGAA ACCCATGCAC TGTGAATCTA GGATTATTC AAAACAAAGG TTTACAGAAA CATCCAAGGA 2100  
 CAGGGCTGAA GTGCCCTCCG GCAAGGGCAG GGCAGGCAGG AGTGATTTTA TTTAGCTATT TTATTTTATT 2170  
 TACTTACTTT CTGAGACAGA GTTATGCTCT TGTGGCCAG GCTGGAGTGC AGCGGCATGA TCTGGCTCA 2240  
 CTGCACCTC CGTCTCTGG GTTCAAGCAA TTCTCGTGC TCAGCCTCCC AAGTAGCTGG GATTTAGGC 2310  
 GTGCACCAAC ACACCCGGCT AATTTGTAT TTTTAGTAGA GATGGGCTT CACCATGTTG GTCAAGCTGA 2380  
 TCTCAAAATC CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC 2450  
 CACTGCACCT GGCCTATTTA ACCATTTTAA AACTTCCCTG GCCTCAAGTC ACACCCACTG GTAAGGAGTT 2520  
 CATGGAGTTC AATTTCCCTT TTAAGTGAAG GTTACCTCC TTTGATATT TCTGTAATTC TTCGTAGACT 2590  
 GGGGATACAC CGTCTCTGA CATATTACA GTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG 2660  
 GGGCAGCTGG GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAGC CATGATGAGA 2730  
 ATCAGGGGCG AAGTGTGGAC ACTGCTCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA 2800  
 AAGTCCATCC CTCCTACTCT ACTGGGATTG AGCCCTTCC CTATCCCCC CCAGGGGCGC AGGAGTCTCT 2870  
 CTCACCTCTG TGGAGGAAGG AATGATACTT TGTATTTTT CACTGCTGGT ACTGAATCCA CTGTTTCAAT 2940  
 TGTGTGTTG TTTGTTTGT TTTGAGAGGC GGTTCACCTC TGTGTGCTCA GGCTGGAGGG AGTGCAATGG 3010  
 CGCGATCTTG GCTTACTGCA GCCTCTGCTT CCCAGGTTCA AGTGATTCTC CTGCTTCCCG CTCCTATTG 3080  
 GCTGGGATTA CAGGCACCCG CCACCATGCC CAGCTAATTT TTTGTATTT TAGTAGAGAC GGGGGTGGGT 3150  
 GGGGTTACAC ATGTTGGCCA GGCTGGTCTC GAATTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCCTCT 3220  
 AAGTGTCTGG GATTACAGGT GTGAGCCACC ATGCCAGCT CAGAAATTAC TCTGTTTAGA AACATCTGGG 3290  
 TCTGAGGTAG GAAGCTCACC CCACCTAAGT GTTGTGGTGT TTTAAGCCAA TGATAGATT TTTTATTGT 3360  
 TGTTAGAACA CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTCAAGG ACACACTAAC 3430  
 TGCACCCATA ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCTCTG 3500  
 CCATGCACAT GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCT TCTCTTCCCT CTTTAAAT 3570  
 TGTGTTTTCT ATGTTGGCTT CTCTGCAGAG AACAGTGA AGCTACAAC TAACTTTTGT TGGAAACAA 3640  
 TTTCCAAACC GCCCTTTGC CCTAGTGGCA GAGACAATTC ACAAACACAG CCCTTTAAAA AGGCTTAGGG 3710  
 ATCACTAAGG GGATTCTAG AAGAGCGACC TGTAACTCTA AGTATTACA AGACGAGGCT AACCTCCAGC 3780  
 GAGCGTGACA GCCCAGGGAG GGTGCGAGGC CTGTTCAAT GCTAGCTCCA TAAATAAAGC AATTTCTCTC 3850  
 GGCAGTTTCT GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTGTGAGC ATTTAGTGT TTGCCGACCT 3920  
 CAGCTACAGC ATCCCTGCAA GGCCTCGGGA GACCCAGAAG TTTCTCGCC CCTTAGATCC AAACCTGAGC 3990  
 AACCCGGAGT CTGGATTCTT GGGAACTCTT CAGCTGTCTT GCGGTTGTG CCGGGCCCCA GGTCTGGAGG 4060  
 GGACCACTGG CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CCGGCCCTCT AGCTCTGAG TCCGAGGCTT 4130  
 GGAGCCAGGT GCCTGGACCC CGAGGCTGCC CTCCACCCTG TGCGGGCGGG ATGTGACCAG ATGTTGGCCT 4200  
 CATCTGCCAG ACAGAGTGCC GGGGCCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG 4270  
 GCCACAGGA GCGCTGGCT CCATTCCCA CCGTTTCTG ACGGGACCGC CCGGTGGGT GATTAAACA 4340  
 TTTGGGGTGG TTTGCTCATG GTGGGACCC CTGCGCGCT GAGAACCCTG AAAGAGAAAT GACGGGCTG 4410  
 TGTCAAGGAG CCCAGTCCG GGGGAAGTGT TGCAGGAGG CACTCCGGGA GGTCCCGCGT GCCCGTCCAG 4480  
 GGAGCAATGC GTCTCGGGT TCGTCCCCG CCGCTCTAC CGCCTCCGT CTTCCCTTC ACGTCCGGA 4550  
 TTCGTGTGTC CCGGAGCCG ACGCCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA 4620  
 GCGGCCAAAG GGTGCCCGCA CGCACCTGTT CCCAGGGCCT CCACATCATG GCCCTTCCCT CGGGTACCC 4690

Fig. 4 (Fortsetzung)

```
CACAGCCTAG GCCGATTGGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA 4760
GCGGCGCGCG GCGGGGAAG CGCGGCCAG ACCCCCGGGT CCGCCCGGAG CAGCTGCGCT GTCGGGGCCA 4830
GGCCGGGCTC CCAGTGGATT CGCGGGCACA GACGCCAGG ACCGCGCTCC CCACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCCGTCCC 4970
GACCCCTCCC GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCC CCCCTTCCTT TCCGCGGGCC 5040
CGCCCTCTCC TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC 5110
GGCCACCCC GCGATG 5126
```



Fig. 5

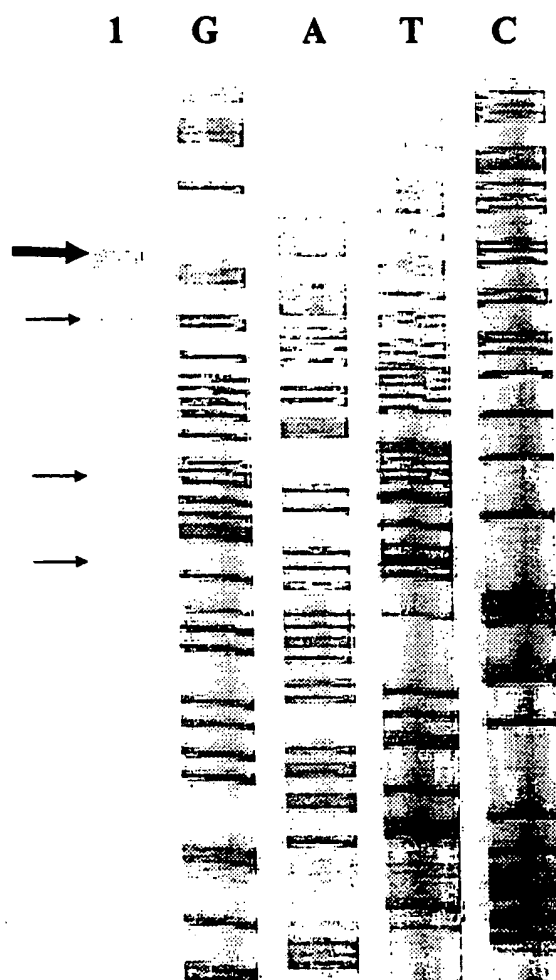


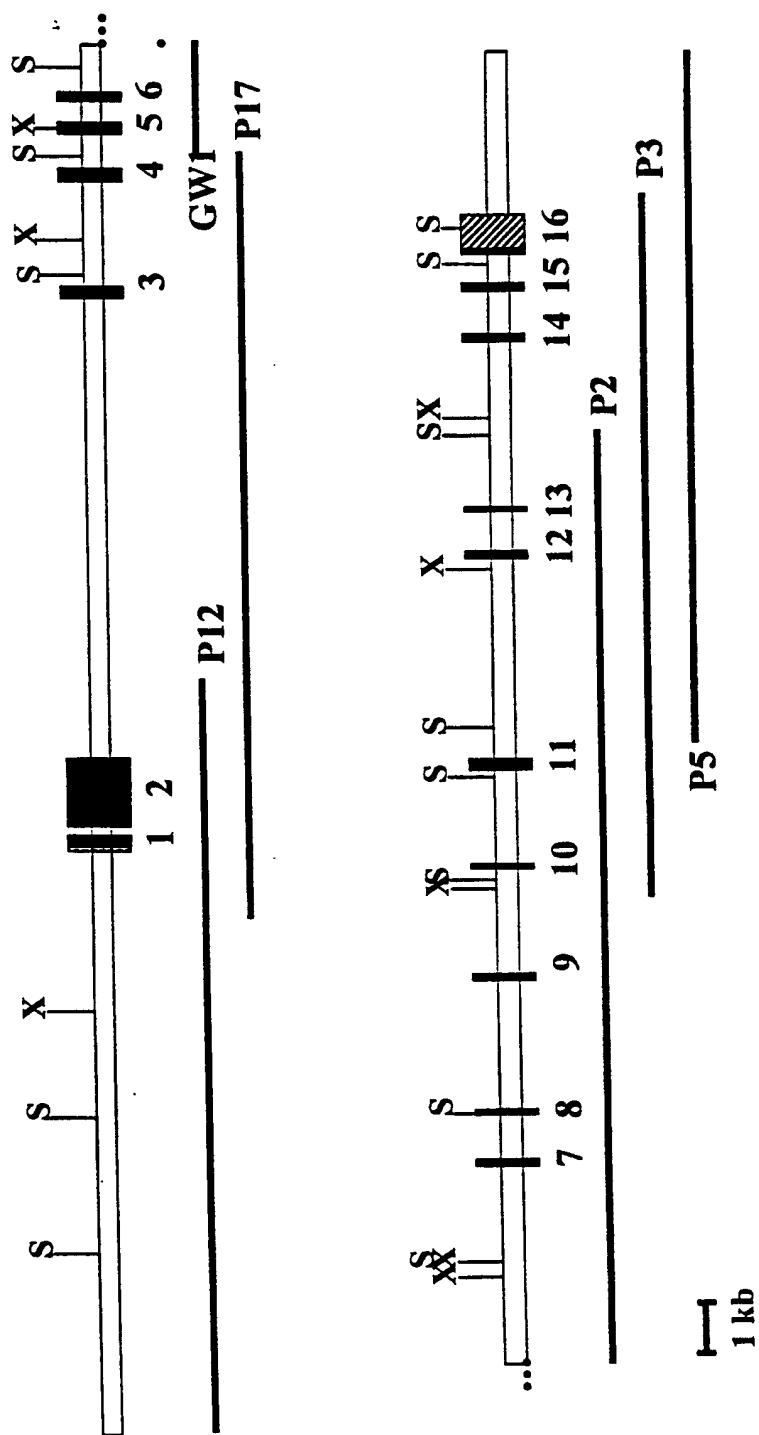
Fig. 6

```

GTTTCAGGCA GCGCTGCGTC CTGCTGCGCA CGTGGGAAGC CCTGGCCCCG GCCACCCCCG CGATGCCCGG 70
CGCTCCCCCG TGCCGAGCCG TGCCTCCCTT GCTGCGCAGC CACTACCGCG AGGTGCTGCC GCTGGCCACG 140
TTCGTGCGGC GCCTGGGGCC CCAGGGCTGG CGGCTGGTGC AGCGCGGGGA CCCGGCGGCT TCCCGCGCCG 210
TGGTGGCCCA GTGCTGGTGG TGCCTGCCCT GGGACGCACG GCCGCCCCCC GCGCGCCCTT CTTCCGCCCA 280
GGTGTCTGCG CTGAAGGAGC TGGTGGCCCG AGTGTGTCAG AGGCTGTGCG AGCGCGGGCG GAAGAACGTG 350
CTGGCCTTCG GCTTCGCGCT GCTGGACGGG GCCCGCGGGG GCGCCCCCGA GGCCTTCACC ACCAGCGTGC 420
GCAGCTACCT GCCCAACACG GTGACCGACG CACTGCGGGG GAGCGGGGCG TGGGGGCTGC TGTGCGCCG 490
CGTGGGCGAC GACGTGCTGG TTCACCTGCT GGCACGCTGC GCGCTCTTTG TGTGGTGGC TCCAGCTGTC 560
GCCTACCAGG TGTGCGGGCC GCCGCTGTAC CAGCTCGGCG CTGCCACTCA GCGCGGGCCC CCGCCACACG 630
CTAGTGGACC CCGAAGCGGT CTGGGATGCG AACGGGCTG GAACCATAGC GTCAGGGAGG CCGGGGTCCC 700
CCTGGGGCTG CCAGCCCCGG GTGCGAGGAG GCGCGGGGGC AGTGCCAGCC GAAGTCTGCC GTTGCCCAAG 770
AGGCGCCAGC GTGGCGCTGC CCTTGAGCCG GAGCGGACGC CCGTTGGGCA GGGGTCTCTG GCCCACCCGG 840
GCAGGACGCG TGGACCGAGT GACCGTGGTT TCTGTGTGGT GTCACCTGCC AGACCCCGCG AAGAAGCCAC 910
CTCTTTGGAG GGTGCGCTCT CTGGCAGCGG CCACTCCACG CCATCCGTGG GCGGCCAGCG CCACCGGGCG 980
CCCCCATCCA CATCGCGGCC ACCACGTCCC TGGACACGCG CTGTCCCCC GGTGTACGCC GAGACCAAGC 1050
ACTTCTCTTA CTCTCAGGCG GACAAGGAGC AGCTCGGGCC CTCTTCTTA CTCAGCTCTC TGAGGCCACG 1120
CCTGACTGGC GCTCGGAGGC TCGTGGAGAC CATCTTCTG GGTTCAGGC CCTGGATGCC AGGGACTCCC 1190
CGCAGGTTGC CCCGCTGCC CCAGCGCTAC TGGCAATGC GCGCCCTGTT TCTGGAGCTG CTTGGGAACC 1260
ACGCGCAGTG CCCCTACGGG GTGCTCTCA AGACGCACTG CCGCTGCGA GCTGCGGTCA CCCCAGCAGC 1330
CGGTGTCTGT GCGCGGAGA AGCCCCAGGG CTCTGTGGC GCGCCGAGG AGGAGGACAC AGACCCCGT 1400
CGCTGTGGTGC AGCTGCTCCG CCAGCACAGC AGCCCTGGC AGGTGTACGG CTTCGTGCGG GCGTGCCTGC 1470
GCGCGCTGGT GCGCCAGGC CTCTGGGGCT CCAGGCACAA CGAACGCGCG TTCCTCAGGA ACACCAAGAA 1540
GTTTCATCTC CTGGGAAGC ATGCCAAGCT CTCGCTGCAG GAGCTGACGT GGAAGATGAG CGTGGCGGAC 1610
TGCCTTTGGC TCGCAGGAG CCCAGGGGTT GGCTGTGTTT CCGCCGACGA GCACCGCTCG CTGTAGGAGA 1680
TCCTGGCCAA GTTCTGTCAC TGGCTGATGA GTGTGTACGT CGTCGAGCTG CTCAGGTCTT TCTTTTATGT 1750
CACGAGAGCC ACGTTTCAAA AGAACAGGCT CTTTTTTAC CGGAAGAGTG TCTGGAGCAA GTTGCAAGC 1820
ATTGAATCA GACAGCACTT GAAGAGGTTG CAGCTCGGGG AGCTGTCCGA AGCAGAGGTC AGGCAGCATC 1890
GGGAAGCCAG GCGCGCCCTG CTGACGTCCA GACTCCGCTT CATCCCCAAG CCTGACGGGC TCGGGCCGAT 1960
TGTGAACATG GACTACGTCG TGGGAGCCAG AACGTTCCGC AGAGAAAAGA GGGCCGAGCG TCTCACTCTG 2030
AGGGTGAAGG CACTGTTTCA CGTGCTCAAC TACGAGCGGG CCGCGCGCCC GCGCCTCTG GCGCCTCTG 2100
CCTGTGCTAC GCGCATATC CACAGGGGCT GCGCACCTT CGTGCTGCGT GTGCGGGGCC AGGACCCGCG 2170
GCCTGAGCTG TACTTTGTCA AGGTGGATGT GACGGGCGCG TACGACACCA TCCCCCAGGA CAGGCTCAGC 2240
GAGGTCACTG CCAGCATCAT CAACCCCGAG AACACGTACT GCGTGCCTCG GTATGCCGTC GTCCAGAAGG 2310
CCGCCCCATG GCACGTCCCG AAGGCTTCA AGAGCCACGT CTCTACCTTG ACAGACCTCC AGCCGTACAT 2380
GCGACAGTTC GTGGCTCACC TGCAGGAGAC CAGCCCGCTG AGGGATGCCG TCGTCACTGA GCAGAGCTCC 2450
TCCCTGAATG AGGCCAGCAG TGGCCTCTTC GACGTCTTCC TACGCTTCAT GTGCCACCA GCGGTGCGCA 2520
TCAGGGGCAA GTCTACGTC CAGTGCCAGG GGATCCCGCA GGGCTCCATC CTCTCCACG CTCTGTCAG 2590
CCTGTGCTAC GCGCATATG AGAACAGCT GTTTGCGGGG ATTGCGCGGG AGGGGCTGCT CCGCGTTTG 2660
GTGGATGATT TCTGTGTGGT GACACCTCAC CTCACCCACG CGAAAACCTT CCTCAGGACC CTGGTCCGAG 2730
GTGTCCCTGA GTATGGCTGC GTGGTGAAT TCGGGAAGAC AGTGGTGAAC TTCCCTGTAG AAGACGAGGC 2800
CCTGGGTGCG ACGGCTTTTG TTCAGATGCC GCGCCACGGC CTATTCCTT GGTGCGGCTT GCTGCTGGAT 2870
ACCCGAGCCC TGGAGGTGCA GAGCGACTAC TCCAGCTATG CCGGACCTC CATCAGAGCC AGTCTCACT 2940
TCAACCGCGG CTTCAGGCTT GGGAGGAACA TGCCTGCGAA ACTCTTTGG GTCTTGCGGC TGAAGTGTCA 3010
CAGCCTGTTT CTGGATTGCG AGGTGAACAG CCTCCAGACG GTGTGCACCA ACATCTACAA GATCCTCTCT 3080
CTGCAGGCGT ACAGGTTTCA CGCATGTGTG CTGCAGCTCC CATTTTCATCA GCAAGTTTGG AAGAACCCCA 3150
CATTTTCTCT GCGCGTCATC TCTGACACGG CCTCCCTCTG CTACTCCATC CTGAAAGCCA AGAACGCGG 3220
GATGTCTGCT GGGGCCAAGG GCGCGCGCGG CCTCTGCGCC TCCGAGGCGG TGCACTGGCT GTGCCACCA 3290
GCATTCTCTG TCAAGCTGAC TCGACACCGT GTCACCTACG TGGCACTCTT GGGGTCACTC AGGACAGCCC 3360
AGACGCACTG GAGTCGGAAG CTCCCGGGGA CGACGCTGAC TGCCTGGAG GCGCGAGCCA ACCCGGCACT 3430
GCGCTCAGAC TTCAGACCA TCCTGGAGTG ATGGCCACCC GCGCCACAGC AGGCGGAGG CAGACCCAG 3500
CAGCCCTGTC ACGCGGGCT CTACGTCCA GGGAGGGAGG GCGCGGCCAC ACCCAGGCCC GCACCGCTGG 3570
GAGTCTGAGG CCTGAGTGAG TGTTTGGCGG AGGCTGCTAT GTCCGGCTGA AGGCTGAGTG TCCGGCTGAG 3640
GCCTGAGCGA GTGTCCAGCC AAGGGCTGAG TGTCCAGCAC ACCTGCGCTC TTCACTTCCC CACAGGCTGG 3710
CGCTCGGCTC CACCCAGGG CCAGCTTTTC CTCACCAAGG GCGCGGCTTC CACTCCCCAC ATAGGAATAG 3780
TCCATCCCCA GATTGCGCAT TGTTCACCCC TCGCCCTGCC CTCCTTTGCC TTCCACCCCC ACCATCCAGG 3850
TGGAGACCCCT GAGAAGGACC CTGGGAGCTC TGGGAATTTG GAGTGACCAA AGGTGTGCCC TGTACACAGG 3920
CGAGGACCCCT GCACCTGGAT GGGGTCTCCCT GTGGGTCAAA TTGGGGGAG GTGCTGTGGG AGTAAATATC 3990
TGAATATATG AGTTTTCAG TTTTGAATAA AAAAAAAAAA AAAAAAAAAA AA 4042

```

Fig. 7



**Fig. 8A**

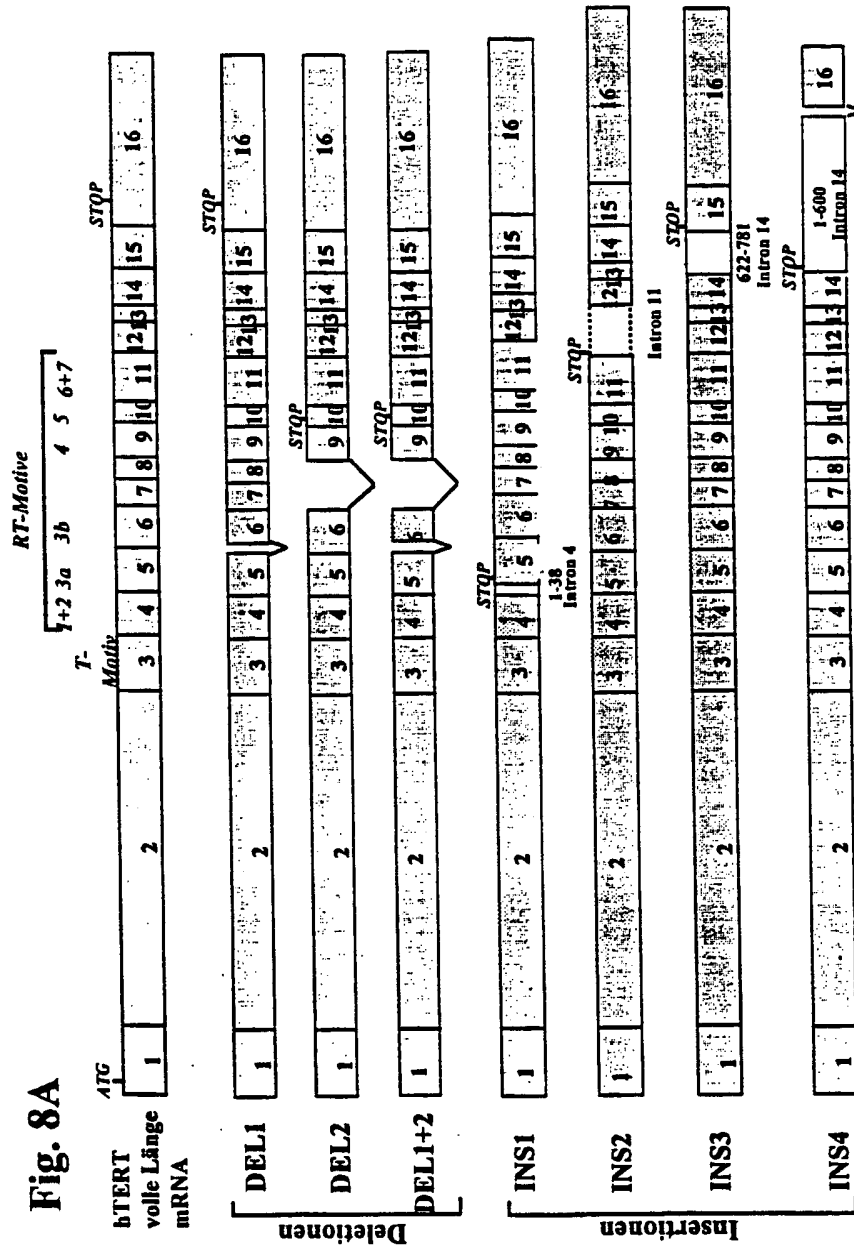
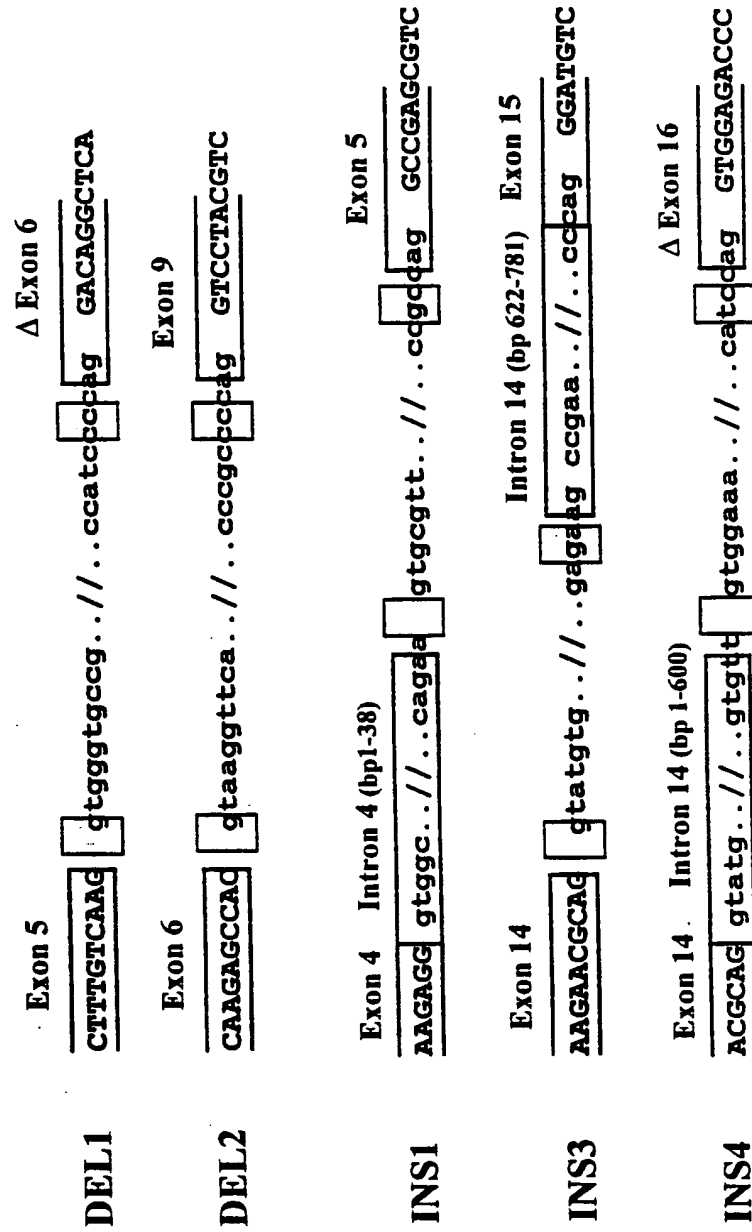


Fig. 8B



**Fig. 9**

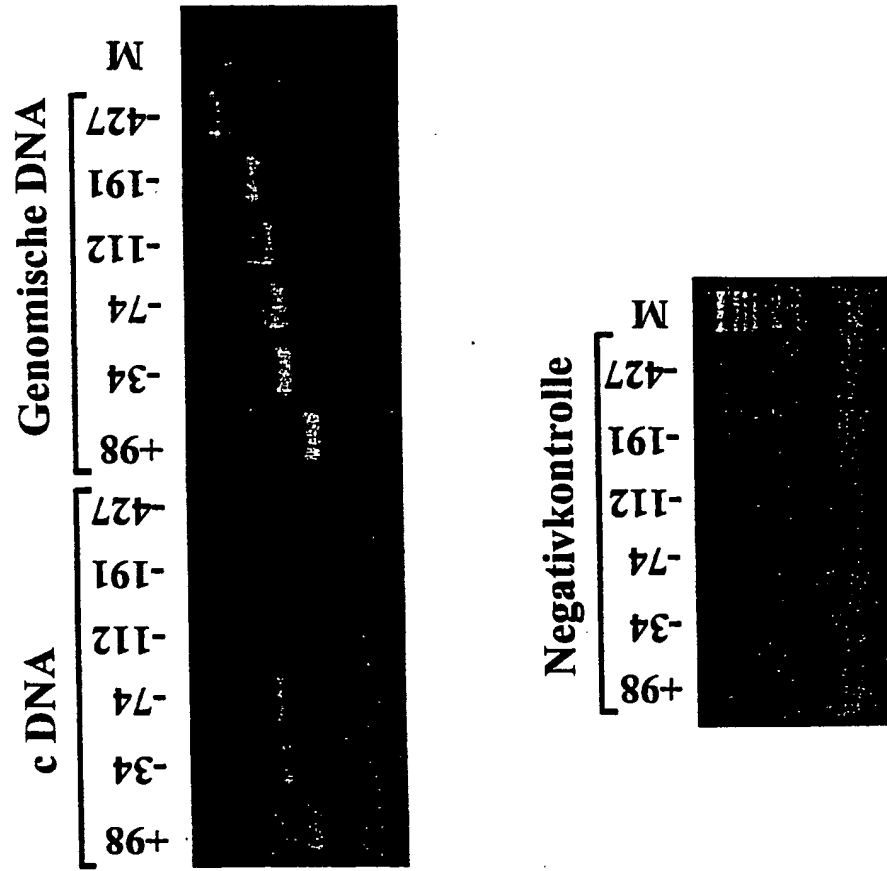


Fig. 10

ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204  
 ATGAGACCCT GTCTCAAAAA AAAAAAAAAA AATTGAAATA ATATAAAGCA TCTTCTCTGG CCACAGTGGA -11134  
 AAAAAACCAG AAATCAACAA CAAGAGGAAT TTTGAAACT ATACAAACAC ATGAAATTA AACAAATATAC -11064  
 TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTAAGAGGA AATTGAAAAA TTTATTTAAG CAAATGATAA -10994  
 CGGAAACATA ACCTCTCAAA ACCCACGGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA -10924  
 AGGAGCTACA TCAAAAAAGT AGAAAAGCCA GGCGCAGTGG CTCATGCCTG TAATCCCGAG ACTTTGGGAG -10854  
 GCCAAGGGGG GCAGATCGCC TGAGGTGAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTCG -10784  
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCAGATG CCTGTAATCC CAGTACTCG GGAGGCTGAG -10714  
 GCAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGGAT TGCGCCATTG GACTCCAGCC -10644  
 TGGGTAACAA GAGTGAAACC CTGTCTCAAG AAAAAAAAAA AAGTAGAAAA ACTTAAAAAT ACAACCTAAT -10574  
 GATGCACCTT AAAGAAGTAG AAAAGCAAGA GCAAACTAAA CCTAAAAATT GTAAAAAGAA AGAATTAATA -10504  
 AAGATCAGAG CAGAAATAAA TGAAGCTGAA AGATAACAAT ACAAAGATC AACAAAATTA AAAGTTGGTT -10434  
 TTTTGAAGAG ATAAACAAAA TTGACAAACC TTTGCCAGA CTAAGAAAAA AGGAAAGAA AGCTAAATAA -10364  
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACTGATAC CACAGAAAT CAAAGGATCA CTAGAGGCTA -10294  
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCCTAGA TGCATACAAC -10224  
 CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAACCA GACCAATAAC AATAATGGGA TTAAGCCAT -10154  
 AATAAAAGT CTCCTAGCAA AGAGAAGCCC AGGACCCAAT GGCTTCCCTG CTGGATTTTA CCAATCATTT -10084  
 AAAGAAGAGT GAATTCATCT CTAATCTCAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACTCAT -10014  
 TCTACATGGC CAGTATTACC CTGATTCCAA AACCAGACAA AAACACATCA AAAACAAACA AACAAAAAAA -9944  
 CAGAAAGAAA GAAACTACA GGCCAAATATC CCTGATGAAT ACTGATACAA AAATCCTCAA CAAACACTA -9874  
 GCRAACCAAA TTAACAACA CCTTCGAAAG ATCATTCAAT GTGATCAAGT GGGATTTAT CCAGGGATGG -9804  
 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAAACTA -9734  
 TATGATTATT TCACTTTATG CAGAAAAAGC ATTTGATAAA ATTTCTGCACC CTTCTAGATA AAAACCCCTCA -9664  
 AAAAACCCAGG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TGCGATCCCA GCACCTCTGG -9594  
 AGGCCAAGGT GGGATGATTG CTTGGGCCCC GGAGTTTGAG ACTAGCCTGG GCAACAAAAA GAGACCTGGT -9524  
 CTACAAAAAA CTTTTTAA AAATTAGCCA GGCGATGGG CATATGCCTG TAGTCCCAGC TAGTCTGGAG -9454  
 GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCAATGAGC CATGAAACATG TCACTGTACT -9384  
 CCAGCCTAGA CAACGAGACA AGACCCCACT GAATAAGAGG AAGGAGAAGG AGAAGGGAGA AGGGAGGGAG -9314  
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG -9244  
 AAGAAACATA TTTCAACATA ATAAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAAGCTGA -9174  
 AGCCTTTCCT CTAAGATCTG GAAAATGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTACTAGA -9104  
 AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAAGGA AGAAGTCAA -9034  
 TTATCTGTT TGCAGATGAT ATGATCTTAT ATCTGGAAGA GACTTAAGAC ACCACTAAAA AACTATTAGA -8964  
 GCTGAAATTT GGTACAGCAG GATACAAAA CAATGTACAA AAATCAGTAG TATTTCTATA TTCCAACAGC -8894  
 AAACAATCTG AAAAAAGAAC CAAAAAGCA GCTACAAATA AAATTAACA GCTAGGAATT AACCAAAGAA -8824  
 GTGAAAGAT TCTACAATGA AAATAATAA ATGTTGATAA AAGAAATTGA AGAGGGCACA AAAAAAGAA -8754  
 AGATATTCCA TGTTCTAGTA TTGGAAGAAT AAATCTGTT AAAATGTCCA TACTACCAA AGCAATTAC -8684  
 AAATTCATG CAATCCCTAT TAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTCTAAGAT -8614  
 TTGTACAGAA CCACAAAGA CCCAGAATAG CCAAGCTAT CCTGACCAA AAGAACAAAA CTGGAAGCAT -8544  
 CACATTACCT GACTTCAAAT TATATACAA AGCTATAGTA ACCCAAATA CATGGTATG GCATAAAAAAC -8474  
 AGATGAGACA TGGACCAGAG GAACAGATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAACCTA -8404  
 TTTTGTACAA AGGTGCCAAG AACATACTTT GGGGAAAAGA TAATCTCTC AATAAATGGT GCTGGAGGAA -8334  
 CTGGATATCC ATATGCAAAA TAACAATCT AGAAGCTCTG CTCTCACCAT ATACAAAAGC AAATCAAAAT -8264  
 GGATGAAAGG CTTAAATCTA AAACCTCAA CTTTGCAACT ACTAAAAGAA AACACCGGAG AAACCTCTCA -8194  
 GGACATTGGA GTGGGCAAG ACTTCTTGAG TAATCCCTG CAGGCACAGG CAACCAAAGC AAAACAGAGC -8124  
 AAATGGGATC ATATCAAGTT AAAAAGCTTC TGCCAGCAA AGGAAACAAT CAACAAAGAG AAGAGACAAC -8054  
 CCACAGAATG GGAGAATATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATAACCAGTA TATATAAGGA -7984  
 GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAT AAGCAAAGA TCTGGGTAGA -7914  
 CATTTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAAAAATGTG CTCAACACCA CTGATCATCA -7844  
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTCAAAGAGC -7774  
 AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGA -7704  
 TTGCTACCAC TATGGAGAAC AGTTTGAAG TTTCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC -7634  
 CCATTGCTAG GTATATACTC CAAAAAGGG AATCAGTGTA TCAACAAGCT ATCTCCACTC CCACATTTAC -7564  
 TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCAT CAACAGACGA ATGGAAAAAG -7494  
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA -7424  
 CAGCATGGGG GGCAGTGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT -7354  
 CTCCTTACT TGTGGGAGCA AAAATTAATA CAATTGACAT AGAAATAGAG GAGAATGGTG GTTCTAGAGG -7284  
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTTAA AATAACTAAA AGAGTATAAT -7214  
 TGGGTTGTTT GTAACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCATTTAC CTGATGTGA -7144  
 TTATTACACA TTGTATGCCT GTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCCTA CTTATTAAA -7074  
 AATTAATAAT TTAATGGCCA GGCACGGTGG CTCATGTCCG TAATCCAGC ACTTTGGGAG GCCGAGGCGG -7004  
 GTGGATCACC TGAGGTGAGG AGTTTGAAAC CAGTCTGGCC ACCATGATGA AACCTGTCT CTAATAAGA -6934  
 TACAAAAAT AGCCAGGCGT GGTGGCAGAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA -6864  
 TTGCTTGAAC CTGGGAGGCG GAGGTGTCAG TGAGCCGAGA TCATGCCACT GCACGTGACA CTGGGTGACA -6794  
 GAGCAAGACT CCATCTCAA ACAAAACAA AAAAAAGAG ATTAATAATTG TAATTTTAT GTACCGTATA -6724  
 AATATATACT CTAATATATT AGAAGTTAAA AATTAACA ATTATAAAG GTAATTAACC ACTTAATCTA -6654  
 AAATAAGAAC AATGTATGTG GGGTTTCTAG CTTCTGAAGA AGTAAAGTT ATGGCCACGA TGGCAGAAAT -6584

ERSATZBLATT (REGEL 26)

Fig. 10

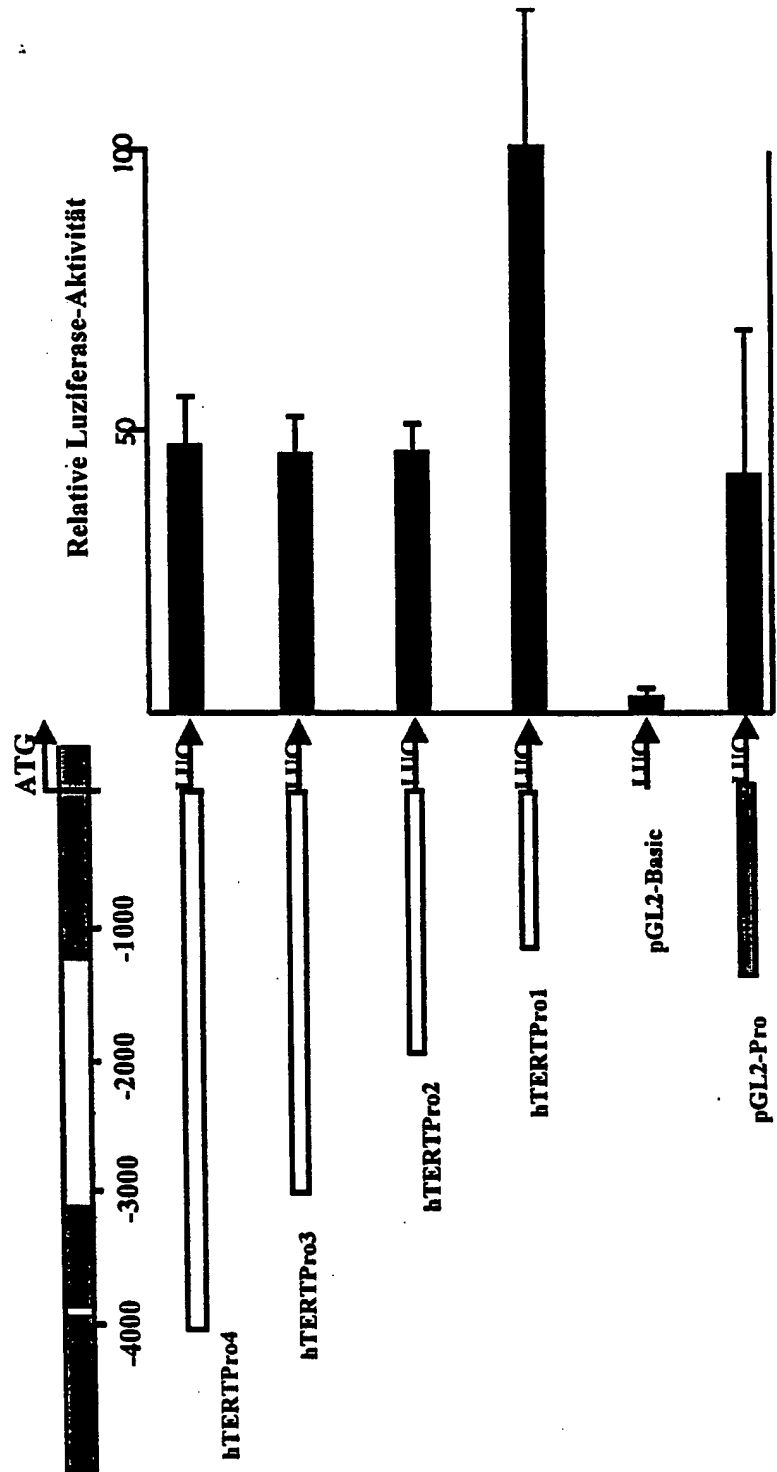
GTGAGGAGGG AACAGTGGAA GTTACTGTGT TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAACC -6514  
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAAGG -6444  
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374  
 CACCGTCCTC TCATTACAGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG -6304  
 GTTAAACTTA ATCTGTATGA ATCCTGAAAC GAAAAATGGT GGTGATTTCC TCCAGAAGAA TTAGAGTACC -6234  
 TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164  
 TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094  
 CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAGATTG CTCTGGATAC CATCTGGAAA -6024  
 AGGCGGCGAG CGGGAATGCA AGGAGTCAGA AGCCTCCTGC TCAAACCCAG GCCAGCAGCT ATGGCGCCCA -5954  
 CCGGGCGCTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTACCTGA -5884  
 AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCTTCTT TAAACAGAA AGTCATGGAA -5814  
 GCACCCCTCT CAAGGGAAAA CCAGACGCCC GCTCTGCGGT CATTTACCTC TTTCTCTCTT CCCTCTCTTG -5744  
 CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCCTGGGA GCTTCTCCGA GCCCGTGTG AGGACCTCT -5674  
 TGCAAAAGGGC TCCACAGACC CCGCCCTGGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG -5604  
 TGGCTGGGGG CGGACAGCGA CCGCGGGATT CAAAGACTTA ATTCCATGAG TAAATTCAC CTTTCCACAT -5534  
 CCGAATGGAT TTGGAATTTA TCTTAATATT TTCTTAAATT TCATCAATA ACATTCCAGGA CTGCAGAAAT -5464  
 CCAAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394  
 ATTTTTCGCC CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG -5324  
 AGAGGCCCTG GCGGCAGGGC TATGAGCAGC GCAGGGCCAC CGGGGAGAGA GTCCCCGCC TGGGAGGCTG -5254  
 ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGCCACA TTGGGCAACG CGAAGGCGGC CACGCTGCGT -5184  
 GTGACTCAGG ACCCCATACC GGCTTCCTGG GCCACCCAC ACTAACCAG GAAGTACAGG AGCTCTGAAC -5114  
 CCGTGGAAC GAACATGACC CTTGCCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA TGGTGTGCA -5044  
 GGAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTATT CATCTTACC -4974  
 CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAGCCATGA CAAAACCTAG TACAAACACC -4904  
 ACTCTTTTAC TAGGCCACA GAGCAGGSC CACACCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834  
 GCTTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTCC TCTAGACTAG TAGACCCTGG -4764  
 CAGGCACTCC CCCAGATTCT AGGGCTGCTG TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC -4694  
 CTGGGGTGCC ACACGTAGGC CAGCCCTGTC TCCACACCCT CCGCTCCAG GCCTCAGCTT CTCCAGCAGC -4624  
 TTCTTAAACC CTGGGTGGGC CGTGTCCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGTCTCAGCG -4554  
 ACGTAGCTCG CACGGTTCCT CCTCACATGG GGTGTCTGTC TCCTTCCCA ACACCTCAT GCGTTGAAGG -4484  
 GAGGAGATTC TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GATGCAGGTT -4414  
 CCTGGCTGCT GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCCGTCTC CTGTCTCTG CCGGGGCTG -4344  
 CCGGTGTGTT CTTCTGTTTC TGTGCTCCTT TCCACGTCCA GCGCTGTG TCTCTGCCG CTAGGGTCTC -4274  
 GGGGTTTTTA TAGGCATAGG ACGGGGGCGT GGTGGGCCAG GCGCTCTTG GGAATGCAA CATTTGGGTG -4204  
 TGAAAGTAGG AGTGCTGTC CTCACCTAGG TCCACGGGCA CAGGCCTGGG GATGGAGCCC CCGCCAGGGA -4134  
 CCGGCCCTTC TCTGCCAGC ACTTTCCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT TTCCACAAGC -4064  
 ACTAAGCATC CTCTTCCCAA AAGACCCAGC ATTGGCACCC CTGGACATTT GCCCCACAGC CTGGGAATT -3994

*~MVC*  
 CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCG CTGTTTTATT TTAATAGCTA -3924  
 CAAAGCAGGG AAATCCCTGC TAAATGTCC TTTAACAAAC TGGTTAAACA AACGGGTCCA TCCGCACGGT -3854  
 GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTAACGT -3784  
 GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAG AAGAATTTT ACCCATGGC -3714  
 AGGGGAGTGG TTAGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCACGCA CCTTTTACTA -3644  
 AAGCCAGTTT CCTGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG -3574  
 GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCTGGGCAGG ATAATGCTCT AGAGATGCC -3504  
 ACGTCTGAT TCCCCCAAAC CTGTGGACAG AACCCTCCCG GCCCCAGGCG CTTTGCAAGT GTGATCTCCG -3434  
 TGAGGACCCT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGGCCGAAA GTAATCCAGG GGTTCCTGGA -3364  
 AGAGGCGGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGCTGAAAAG -3294  
 GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCCTCCAG AAGCTGAAA AAGCGGGGAA GGGACCTCC -3224  
 ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCTTAGCCCC ACCAGGGCCC ATCGTGAGC TCCGGCTCC -3154  
 GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGG ATTTGCAGAA GCAACAGGAA -3084  
 ACCCATGCAC TGTGAATCTA GGATTATTTC AAAACAAAGG TTTACAGAAA CATCCAAGGA CAGGGCTGAA -3014  
 GTGCCCTCCG GCAAGGGCAG GGCAGGCACG AGTGATTTTA TTTAGCTATT TTATTTTATT TACTTACTTT -2944  
 CTGAGACAGA GTTATGCTCT TGTGCCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874  
 CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTACAGC GTGCACCACC -2804  
 ACACCCGGCT AATTTTGTAT TTTTAGTAGA GATGGGCTTT CACCATGTTG GTCAAGCTGA TCTCAAAATC -2734  
 CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664  
 GGCCTATTTA ACCATTTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC -2594  
 AATTTCCCTT TTAAGCAGGA GTTACCCTCC TTTGATATT TCTGTAATTC TTCGTAGACT GGGGATACAC -2524  
 CGTCTCTTGA CATATTCACA GTTCTGTGTA CCACCTGTTA TCCCATGGGA CCCACTGCAG GGGCAGCTGG -2454  
 GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGCGC -2384  
 AAGTGTGGAC ACTGTCTGTA ATCTCAATG CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC -2314  
 CTCCTACTCT ACTGGGATTG AGCCCCCTCC CTATCCCCC CCAGGGGCG AGGAGTTCCT CTCACTCTG -2244  
 TGGAGGAAG AATGATACTT TGTATTTTTT CACTGCTGCT ACTGAATCCA CTGTTTCATT TGTGTTGTG -2174  
 TTTGTTTTGT TTTGAGAGG GGTTCACCT TTGTTGCTGA GGCTGGAGGG AGTGCAATGG CGCGATCTTG -2104  
 GCTTACTGCA GCCTCTGCCT CCCAGTTCA AGTGATTCTC CTGCTTCCG CTCCCATTTG GCTGGGATTA -2034  
 CAGGCACCCG CCACCATGCC CAGCTAATTT TTTGTATTT TAGTAGAGAC GGGGGTGGG GGGGTTCCAC -1964



ATGTTGGCCA GGCCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894  
GATTACAGGT GTGAGCCACC ATGCCACGCT CAGAATTAC TCTGTTTGA AACATCTGGG TCTGAGGTAG -1824  
CAAT-Box  
GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAGGCCAA TATAGAATT TTTTATTGT TGTTAGAACA -1754  
CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC TGCACCCATA -1684  
ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCAATTGA TGCCGGGAGG CGTTTCCTCG CCATGCACAT -1614  
GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCT TCTCTTCCCT CTTTTAAAT TGTGTTTCT -1544  
ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAAC TAACTTTGT TGGAACAAAT TTTCCAAACC -1474  
Spl  
GCCCTTTGC CCTAGTGGCA GAGACAATC ACAAACAG CCCTTTAAAA AGGCTTAGGG ATCACTAAGG -1404  
GGATTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334  
GCCCAGGGAG GTGCGGAGC CTGTTCAAAT GCTAGCTCCA TAAATAAAGC AATTCCTCC GGCAGTTTCT -1264  
GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTGTTAGC ATTTCACTGT TTGCCGACCT CAGCTACAGC -1194  
ATCCCTGCAA GGCTCGGGA GACCCAGAAG TTTCTCGCCC CCTTAGATCC AAACCTGAGC AACCCGGAGT -1124  
CTGGATTCT GGGAAAGTCT CAGCTGTCTT GCGGTTGTGC CGGGGCCCCA GGTCTGGAGG GGACCACTGG -1054  
CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCCTCCT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984  
GCCTGGACCC CGAGGCTGCC CTCACCCCTG TGGGGCGGG ATGTGACCAG ATGTTGGCCT CATCTGCCAG -914  
ACAGAGTGCC GGGGCCCAGG GTCAGGCGG TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG GCCAGCAGGA -844  
CCAC-Box  
GCGCCTGGCT CCATTCCCA CCGTTCTCG ACGGGACCG CCCGTTGGGT GATTACAGA TTTGGGTGG -774  
Spl  
TTTGCTCATG GTGGGACCC CTCGCCCTT GAGAACCTGC AAAGAGAAAT GACGGGCTG TGTCAAGGAG -704  
CCCAAGTCG GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCT GCCCGTCCAG GGAGCAATG -634  
AP-2  
GTCTCGGGT TCGTCCCGAC CCGCTCTAC GCGCCTCCGT CCTCCCTTC ACGTCCGCA TTCGTGGTGC -564  
CCGGAGCCCG ACGCCCCCG TCCGGACCTG GAGGCGACCC TGGGTCTCG GATCAGGCCA GCGGCCAAAG -494  
GGTCGCCGA CGCACCTGT CCCAGGGCCT CCACATCATG GCCCTCCCT CGGGTTACCC CACAGCCTAG -424  
Spl  
GCCGATTGCA CCTCTCTCG CTGGGGCCCT CGCTGGCGTC CTTGCACCCT GGGAGCGCGA GCGGCCGCGG -354  
Spl  
GGCGGGGAAG CGCGGCCAG ACCCCCGGT CCGCGCGGAG CAGCTGCGCT GTCGGGGCCA GGCCGGGCTC -284  
CCAGTGGATT CGCGGGACA GACGCCAGG ACCGCGCTCC CACAGTGGC GAGGGACTGG GGACCCGGGC -214  
c-Myc  
ACCGTCTCG CCCCTTACC TTCCAGCTCC GCCTCTCCG CGCGGACCC GCCCGTCCC GACCCCTCCC -144  
Spl  
GGGTCCCCGG CCCAGCCCC TCCGGGCCCT CCCAGCCCT CCCCTTCTT TCCGCGGCC CGCCCTCTCC -74  
c-Myc  
TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CTTGCTGCG CACGTGGAAG CCTGGCCCC GGCCACCCCC -4  
GCGATG

Fig.: 11



## SEQUENZPROTOKOLL

&lt;110&gt; Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen  
der humanen katalytischen Telomerase-Untereinheit und  
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

&lt;140&gt;

&lt;141&gt;

&lt;160&gt; 20

15 <170> PatentIn Vers. 2.0

&lt;210&gt; 1

&lt;211&gt; 5126

20 <212> DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 1

25 gagctctgaa ccgtggaaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60  
gggtaataaa gtggtgtgca ggaaatggcc atgtaaaatta cactgactctg ctgatgggga 120  
ccgttccttc catcattatt catcttcacc cccaaggact gaatgattcc agcaacttct 180  
tcgggtgtga caagccatga caaaactcag tacaacacc actcttttac taggcccaca 240  
gagcacgggc cacacccctg atataatga agtccaggag agatgaggct gctttcagcc 300  
accaggctgg ggtgacaaca gcggctgaac agtctgttcc tctagactag tagaccctgg 360  
30 caggcactcc cccaaattct agggcctggt tgcctgttcc cgaggcgccc atctgcctcg 420  
gagactcagc ctgggtgtgc aactgaggc cagccctgtc tccacacccc ccgcctccag 480  
gcctcagctt ctccagcagc ttccataacc ctgggtgggc cgtgttccag cgctactgtc 540  
tcacctgtcc cactgtgtct tgcctcagcg acgtagctcg cactggttctt cctcacatgg 600  
ggtgtctgtc tcttcccca aactcacat gcgttgaagg gaggagattc tgcgcctccc 660  
35 agactggctc ctctgagcct gaacctggct cgtggccccc gatgcaggtt cctggcgctc 720  
ggctgcagc tgacctccat ttccaggcgc tcccgtctc ctgtcatctg ccggggcctg 780  
ccgtgtgtt cttctgttct tgcctcctt tccacgtcca gctgcgtgtg tctctgccc 840  
ctagggtctc ggggttttta taggcataag acgggggctg ggtggggcag ggcgctctt 900  
ggaaatgcaa catttgggtg tgaagtagg agtgcctgtc ctacactagg tccacgggca 960  
40 caggcctggg gatggagccc ccgccaggga ccgcccttc tctgcccagc actttcctgc 1020  
ccccctccc ctggaacaca gattggcagt ttccacaagc actaagcacc ctctcccca 1080  
aagaccagc attggcacc ctggacattt gcccacagc cctgggaatt cactgtacta 1140  
cgcacatcat gtacacactc ccgtccacga ccgaccccc ctgttttatt ttaatagcta 1200  
caaagcagg aaatccctgc taaaatgtcc ttaacaaaac tgggttaaca aacgggtcca 1260  
45 tccgcacggt ggacagtcc tcacagtga gaggaacatg ccgtttataa agcctgcagg 1320  
catctcaagg gaattacgct gagtcaaac tgccacctcc atgggatacg tacgcaacat 1380  
gctcaaaaag aaagaatttc accccatggc aggggagtg ttaggggggt taaggacggt 1440  
ggggggcgca gctgggggt actgcacgca ccttttacta aagccagttt cctggttctg 1500  
atggtattgg ctacgttatg ggagactaac cataggggag tggggatggg ggaacccgga 1560  
50 ggctgtgcca tctttgccc gcccagtggt cctgggcagg ataatgctc agagatgccc 1620  
acgtcctgat tccccaaa ctgtggacag aacccgccc gcccagggc ctttgagggt 1680  
gtgatctccg tgaggacct gaggtctggg atccttcggg actacctgca ggcccga 1740  
gtaatccagg ggttctggga agaggcggc aggggggtca gagggggca gcctcaggac 1800  
gatggaggca gtcagtctga ggctgaaaag ggaggagggt cctcgagccc aggcctgcaa 1860  
55 gcgcctccag aagctgga aagcggggaa gggacctcc acggagcctg cagcaggaa 1920  
gcacggctgg cccttagccc accaggccc atcgtggacc tccggcctcc gtgccatagg 1980  
agggcactcg cgctgccct ctacgatgaa gtgtgtggg atttgagaa gcaacaggaa 2040  
acccatgcac tgtgaatcta ggattatttc aaacaaaagg ttacagaaa catccaagga 2100  
cagggtgaa gtgcctccg gcaaggcgag ggcaggcac agtgatttta tttagctatt 2160  
60 ttattttatt tacttacttt ctgagacaga gttatgctt tgttggccag gctggagtgc 2220  
agcggcatga tcttggctca ctgcaacctc cgtctcctgg gttcaagcaa ttctcgtgcc 2280  
tcagcctccc aagtagctgg gatttcaggc gtgcaccacc acacccggct aattttgtat 2340  
tttagtaga gatgggcttt caccatgtg gtcaagctga tcccaaaatc ctgacctcag 2400  
gtgatccgc cactcagcc tccaaaagt ctgggattac aggcattgag cactgcacct 2460

ggcctatttta accatttttaa aacttccctg ggctcaagtc acacccactg gtaaggagtt 2520  
catggagttc aattttccctt ttactcagga gttaccctcc ttgtatattt tctgtaattc 2580  
ttcgtagact ggggatacac cgtctcttga catattcaca gtttctgtga ccacctgtta 2640  
tcccatggga ccactgcag gggcagctgg gaggtgcag gcttcaggtc ccagtgggt 2700  
5 tggcatctgc cagtagaaac ctgatgtaga atcagggcgc aagtgtggac actgtcctga 2760  
atctcaatgt ctcatgtgtg gctgaaacat gtgaaatta aagtcacatc ctctactct 2820  
actgggattg agccccctcc ctatccccc ccaggggcag aggagttcct ctactcctg 2880  
tggagggaag aatgatactt tgttattttt cactgctggg actgaatcca ctgtttcatt 2940  
tgttggtttg tttgttttgt ttgagaggg ggtttcactc ttgttgctca ggctggagg 3000  
10 agtgcaatgg cgcgatcttg gcttactgca gcctctgect cccaggttca agtgattctc 3060  
ctgcttccgc ctcccatttg gctgggatta caggcacccg ccacatgcc cagtaattt 3120  
tttgtatttt tagtagagac ggggtgggtt ggggttcacc atgttgcca ggctgggtctc 3180  
gaactttctga cctcagatga tccacctgcc tctgcctcct aaagtgcctg gattacaggt 3240  
gtgagccacc atgcccagct cagaatttac tctgtttaga aacatctggg tctgaggtag 3300  
15 gaagctcacc ccactcaagt gttgtggtgt tttaagccaa tgatagaatt tttttattgt 3360  
tgtagaaca ctcttgatgt ttacactgt gatgactaag acatcatcag cttttcaaa 3420  
acacactaac tgcaccata atactgggtt gtcttctggg tatcagcaat cttcattgaa 3480  
tgccgggagg cgtttcctcg ccatgcacat ggtgttaatt actccagcat aatcttctgc 3540  
ttccatttct tctcttccct cttttaaaat tgtgttttct atgttggtct ctctgcagag 3600  
20 aaccagtgtg agctacaact taacttttgt tggaaacaaat ttccaaaacc gccctttgc 3660  
cctagtggca gagacaattc acaaacacag ccttttaaaa aggcttaggg atcactaagg 3720  
ggatttctag aagagcgacc tgaatccta agtatttaca agacagggtc aacctccagc 3780  
gagcgtgaca gccaggggag ggtgcgaggg ctgttcaaat gctagctcca taaataaagc 3840  
aatttccctc ggaggtttct gaaagtagga aagggttacct ttaagggtgc gtttggtagc 3900  
25 atttcagtgt ttgcccagct cagctacagc atccctgcaa ggcctcggga gaccagaag 3960  
ttctcgcgc ccttagatcc aaacttgagc aacccggagt ctgatttctt ggggaagcct 4020  
cagctgtcct gcggttgtgc cggggcccca ggtctggagg ggaccagtg cgtgtggct 4080  
tctactgctg ggctggaagt cgggcctcct agctctgcag tccgaggtt ggagccaggt 4140  
gcttggaacc cgaggctgcc ctccacctg tgcgggcggg atgtgaccag atgttggtc 4200  
30 catctgccag acagagtgcc gggggccagg gtcaaggccg ttgtggctgg tgtgaggcgc 4260  
ccggtgcgcg gccagcaggg gcgcctggct ccatttccca cctttctctg acgggaccgc 4320  
cccgtgggtt gattaacaga tttgggttgg ttgtctcatg gtgggaacc ctccgcgcct 4380  
gagaacctgc aaagagaaat gacgggcctg tgtcaaggag cccaagtgc ggggaagtgt 4440  
tgagggagg cactccggga ggtccgcgt gccctccag ggaagcaatgc gtctcgggt 4500  
35 tgcgtcccaag ccgctctac gcgcctcctt cctcccttc acgtccggca ttcgtggtgc 4560  
ccggagcccg acgcccgcg tccggacctg gaggcagccc tgggtctccg gatcaggcca 4620  
gcggccaaaag ggtcgcgcga cgcacctgt cccagggcct ccacatcatg gccctcctt 4680  
cgggttacc cccagcctag gccgattcga cctctctccg ctggggccct cgtggcgct 4740  
cctgcacctt gggagcgcga gcggcgcgcg ggcggggaag cgcggcccg acccccggt 4800  
40 ccgcccggag cagctgcgt gtccggggcca ggccgggctc ccagtggatt cgcgggcaca 4860  
gacgcccagg accgcgtcc ccacgtggcg gagggactgg ggaccgggc acccgtcctg 4920  
ccccttcacc ttccagctcc gcctcctccg cgcggacccc gccctgtccc gacccctccc 4980  
gggtcccccg cccagccccc tccgggcccct cccagccctt ccccttctct tccgcggccc 5040  
gcctctctc tccggcgcg agtttcaggc agcgtctgct cctgctgcgc acgtgggaag 5100  
45 ccttgcccc ggccaccccc gcgatg 5126

&lt;210&gt; 2

&lt;211&gt; 4042

&lt;212&gt; DNA

50 &lt;213&gt; Homo sapiens

&lt;400&gt; 2

gtttcaggca gcgctgcgtc ctgctgcgca cgtgggaagc cctggccccg gccacccccg 60  
cgatgccgcg cgtccccgc tgcgagccg tgcgtccctt gctgcgcagc cactaccgcg 120  
55 aggtgctgcc gctggccacg ttctgtcggc gcctggggcc ccagggtctg cggctggtgc 180  
agcgcgggga cccggcggtt ttccgcgcgc tgggtggcca gtgctggtg tgcgtgcct 240  
gggacgcagc gccgcccccc gccgccccct ccttccgcca ggtgtcctgc ctgaaggagc 300  
tgggtggccc agtgctgcag aggtctgtgc agcgcggcgc gaagaactgt ctggccttcg 360  
60 gcttcgcgtc gctggacggg gcccgcgggg gcccccccg ggccttcacc accagcgtgc 420  
gcagctacct gcccacacg gtgaccgacg cactgcgggg gagcggggcg tgggggctgc 480  
tgctgcgcgc cgtggggcag gacgtgctg ttacactgct ggcacgctgc gcgtctcttg 540  
tgctgggtgg tcccagctgc gcctaccagg tgtgcggggc gccgtgtac cagctcggcg 600  
ctgccactca gggccggccc ccgcccacag ctagtggacc ccgaaggcgt ctgggatgcg 660  
aacgggcctg gaaccatagc gtcagggagg ccggggctcc cctgggcctg ccagccccgg 720  
65 gtgcgaggag gcgcgggggc agtgcagcc gaagtctgct gttgcccag aggccacggc 780

3 / 18

gtggcgctgc cccctgagccg gagcggacgc ccgttgggca ggggtcctgg gccaccaccg 840  
 gcaggacgcg tggaccgagt gaccgtgggt tctgtgtggt gtcacctgcc agaccgcgcg 900  
 aagaagccac ctctttggag ggtgcgctct ctggcacgcg ccactccac ccaccgtgg 960  
 gccgccagca ccacgcgggc ccccatcca catcgcggcc accacgtccc tgggacacgc 1020  
 5 cttgtccccc ggtgtacgcc gagaccaagc acttcctcta ctctcaggc gacaaggagc 1080  
 agctgcggcc ctccctccta ctcagctctc tgaggccag cctgactggc gctcggaggc 1140  
 tegtggagac catctttctg ggttccaggc cctggatgcc agggactccc cgcaggttgc 1200  
 ccgcctgcc ccagcgctac tggcaaatgc ggccctgtt tctggagctg cttgggaacc 1260  
 acgcgcagtg cccctacggg gtgctcctca agacgcactg cccgctgcga gctgcggtca 1320  
 10 cccagcagc cggtgtctgt gcccgggaga agccccagg ctctgtggcg gcccccgagg 1380  
 aggaggacac agacccccgt cgcctgggtg agctgctccg ccagcacagc agccccggc 1440  
 aggtgtacgg ctctgtgcgg gcttgcctgc gccggtggt gcccccaggc ctctggggct 1500  
 ccaggcacaa cgaacgcgc tctctcagga acaccaagaa gttcatctcc ctggggaagc 1560  
 atgccaaagt ctgcgtgcag gagctgactg ggaagatgag cgtgcgggac tgcgcttggc 1620  
 15 tgcgcaggag ccagggggtt ggtgtgttc cggccgcaga gcaccgtctg cgtgaggaga 1680  
 tcttgcccaa gttcctgcac tggctgatga gtgtgactg cgtcagactg ctccaggtctt 1740  
 tcttttatgt caccggagac acgtttcaaa agaacaggct ctttttctac cgggaagagt 1800  
 tctggagcaa gttgcaaaagc attggaatca gacagcactt gaagagggtg cagctgcggg 1860  
 agctgtcggg agcagaggtc aggcagcatc ggggaaggag gcccgccctg ctgactgcca 1920  
 20 gactccgctt catccccaag cctgacgggc tgcggccgat tgtgaacatg gactacgtcg 1980  
 tgggagccag aacgttccgc agagaaaaga gggccgagcg tctcacctcg aggggaagg 2040  
 cactgttcag cgtgctcaac tacgagcggg cgcgcgcgcc cggcctcctg ggcgcctctg 2100  
 tgctgggcct ggacgatata cacagggcct ggcgcacctt cgtgctgcgt gtgcgggccc 2160  
 aggaaccgcc gcctgagctg tactttgtca aggtggatgt gacgggcgcg tacgacacca 2220  
 25 tccccagga caggctcacg gaggtcatcg ccagcatcat caaacccag aacacgtact 2280  
 gctgtgcctg gtatgccgtg gtccagaagg ccgcccattg gcacgtccgc aaggccttca 2340  
 agagccacgt ctctaccttg acagacctcc agccgtacat gcgacagtcc gtggctcacc 2400  
 tgcaaggagc cagccccctg agggatgccc tgcgtatcga gcagagctcc tccctgaatg 2460  
 aggccagcag tggcctcttc gacgtcttcc tacgcttcat gtgccaccac gccgtgcga 2520  
 30 tcaggggcaa gtccctacgtc cagtgccagg ggatcccga gggctccatc ctctccaccg 2580  
 tgctctgcag cctgtgctac ggcgacatgg agaacaagct gtttgcgggg attcggcggg 2640  
 acgggctgct cctgcgtttg gtggatgatt tcttgttggg gacacctcac ctacccaccg 2700  
 cgaaaacctt cctcaggacc ctggtccgag gtgtccctga gtatggctgc gtggtgaact 2760  
 tgcggaagac agtggatgaa tccccgtgag aagacgaggc cctgggtggc acggcttttg 2820  
 35 ttcatatgcc gggccacggc ctattcccc ggtgcggcct gctgctggat acccggaacc 2880  
 tggaggtgca gagcgactac tccagctatg cccggacctc catcagagcc agtctcactt 2940  
 tcaaccgcgg ctccaaggct gggaggaaca tgcgtcgcaa actctttggg gtcttgcggc 3000  
 tgaagtgtca cagctgtttt ctggatttgc aggtgaacag cctccagacg gtgtgcacca 3060  
 40 acatctacaa gatectcctg ctgcaggcgt acagggttca cgcagtgttg ctgcagctcc 3120  
 catttcatca gcaagtgttg aagaacccca catttttctt cgcgctcatc tctgacacgg 3180  
 cctccctctg ctactccatc ctgaaagcca agaacgcagg gatgtcgtg ggggccaagg 3240  
 gcgcgcggcg cctctgccc tccgaggcgg tgcaagtggc gtgccacca gcatctctgc 3300  
 tcaagctgac tcgacaccgt gtcacctacg tgccactcct ggggtcactc aggcagagcc 3360  
 agacgcagct gagtgcgaag ctccccggga gcagctgac tgccctggag gccgcagcca 3420  
 45 acccggcact gccctcagac ttcaagacca tccctggactg atggccaccc gccacagcc 3480  
 agggcgagag cagacaccag cagccctgtc acgcccggct ctacgtccca gggaggaggg 3540  
 ggccgcccac acccaggccc gcaccgctgg gagtctgagg cctgagttag tgtttggccg 3600  
 aggcctgcat gtccggctga aggtgagtg tccggctgag gcctgagcga gtgtccagcc 3660  
 50 aagggtgag tgtccagcac acctgcccgtc ttcaacttccc cacaggctgg cgtcgggtc 3720  
 caccacaggg ccagcttttc ctacaccagga gcccggttcc cactccccac atagggaatg 3780  
 tccatcccca gattcgccat tgttcacccc tcgcccgtgc ctcttttggc ttccaccccc 3840  
 accatccagg tggagaccct gagaaggacc ctgggagctc tgggaatttg gagtgaacaa 3900  
 aggtgtgccc tgtacacagg cgaggacctt gcacctggat ggggtccct gtgggtcaaa 3960  
 55 ttggggggag gtgctgtggg agtaaaatag tgaatatatg agtttttcag ttttgaaaaa 4020  
 aaaaaaaaaa aaaaaaaaaa aa 4042

<210> 3  
 <211> 11276  
 <212> DNA  
 60 <213> Homo sapiens

<400> 3  
 acttgagccc aagagttcaa ggctacggtg agccatgatt gcaacaccac acgcagcct 60  
 65 tgggtgacaga atgagaccct gtctcaaaaa aaaaaaaaaa aattgaaata atataaagca 120  
 tcttctcttg ccacagtgga acaaaaccag aaatcaacaa caagaggaat tttgaaact 180

atcaaaacac atgaaaatta aacaatatac ttctgaatga ccagtgaatc aatgaagaaa 240  
ttaaaaaagga aattgaaaaa ttattttaag caaatgataa cggaaacata acctctcaaa 300  
acccacggta tacagcaaaa gcagtgtctaa gaaggaggtt tatagctata agcagctaca 360  
tcaaaaaagt agaaaagcca ggcgcagtggt ctcctgcctg taatcccagc actttgggag 420  
5 gccaaaggcg gcagatcgcc tgaggtcagg agttcgagac cagcctgacc aacacagaga 480  
aaccttgtcg ctactaaaaa tacaaaatta gctgggcatg gtggcacatg cctgtaatcc 540  
cagctactcg ggaggctgag gcaggataac cgctgaacc caggaggtgg aggttgcggt 600  
gagccgggat tgcgcattg gactccagcc tgggtaacaa gagtgaacc ctgtctcaag 660  
aaaaaaaaa aagtagaaaa acttaaaaat acaacctaat gatgcacct aaagaactag 720  
10 aaaaagcaaga gcaaaactaaa cctaaaattg gtaaaagaaa agaaataata aagatcagag 780  
cagaaataaa tgaactgaa agataacaat acaaaagatc aacaaaatta aaagtgtggt 840  
ttttgaaaag ataaacaaaa ttgacaaacc tttgccaga ctaagaaaaa aggaagaag 900  
acctaataa ataaagtcag agatgaaaaa agagacatta caactgatac cacagaaatt 960  
caaaagatca cttagggcta ctatgagcaa ctgtacacta ataaattgaa aaacctagaa 1020  
15 aaaaatagata aattcctaga tgcatacaac ctaccaagat tgaaccatga agaaatccaa 1080  
agcccaaaa gaccaataac aataatggga ttaaagccat aataaaaaagt ctcttagcaa 1140  
agagaagccc aggaccaat ggcttccctg ctggatttta ccaatcattt aaagaagaat 1200  
gaattccaat cctactcaaa ctattctgaa aaatagagga aagaatactt ccaaactcat 1260  
tctacatggc cagtattacc ctgattccaa aaccagacaa aaacacatca aaaaacaaaca 1320  
20 aacaaaaaaa cagaaagaaa gaaaactaca ggccaatct cctgatgaat actgatacaa 1380  
aaatcctcaa caaaacataa gcaaaccaaa ttaaacacaa ctttcgaaag atcattcatt 1440  
gtgatcaagt gggatttatt ccagggtatg aaggatggtt caacatagtc aaatcaatca 1500  
atgtgatata tcatcccaac aaaatgaagt acaaaaacta tatgattatt tcaatttatg 1560  
cagaaaaagc atttgataaa attctgcacc ctctcatgata aaaaacctca aaaaaccagg 1620  
25 tatacaagaa acatacaggg caggcacagt ggctcacacc tgcgatccca gcaactctgg 1680  
aggccaaggt gggatgatg cttgggccc aaggattttag actagccttg gcaacaaaat 1740  
gagacctggc ctacaaaaaa cttttttaa aaattagcca ggcctgatg catatgcctg 1800  
tagtcccgag tagtctggag gctgaggtgg gagaatcact taagcctagg aggtcgaggc 1860  
tgcagtgaac catgaacatg tcaactgtact ccagcctaga caacagaaca agacccact 1920  
30 gaataaagag aaggagaagg agaagggaga agggagggag aaggaggag gaggagaagg 1980  
aggaggtgga ggagaagtgg aaggggaagg ggaagggaag gaggaagaag aagaacata 2040  
tttcaacata ataaaagccc tatatgacag accgaggtag tattatgagg aaaaactgaa 2100  
agcctttcct ctaagatctg gaaaatgaca agggcccact tccaccactg tgattcaaca 2160  
tagtactaga agtcttagct agagcaatca gataagagaa agaaataaaa ggcattccaaa 2220  
35 ctggaaagga agaaagcaaa ttatcctgtt tgcagatgat atgatcttat atctggaaa 2280  
gacttaagac accactaaaa aactattaga gctgaaattt ggtacagcag gatacaaaat 2340  
caatgtacaa aaatcagtag tatttctata ttccaacagc aaacaatctg aaaaagaaac 2400  
caaaaaagca gctacaaata aaattraaaa gctaggaatt aaccaaagaa gtgaaagatc 2460  
40 tctacaatga aaactataaa atgttgataa aagaaattga agagggcaca aaaaagaaa 2520  
agatattcca tgttcataga ttggaagaat aaatactgtt aaaatgtcca tactaccaa 2580  
agcaatttac aaattcaatg caatccctat taaaatacta atgacgttct tcacagaaat 2640  
agaagaaaca attctaagat ttgtacagaa ccacaaaaga cccagaatag ccaagctat 2700  
cctgaccaa aagaacaaaa ctggaagcat cacattacct gacttcaat tatactaca 2760  
agctatagta acccaacta catggtactg gcataaaaac agatgagaca tggaccagag 2820  
45 gaacagaata gagaatccag aaacaaatcc atgcatctac agtgaactca tttttgacaa 2880  
agggtgccaag aacatacttt ggggaaaaga taatctcttc aataaatggt gctggaggaa 2940  
ctggatatcc atatgcaaaa taacaatact agaactctgt ctctcaccat atacaaaagc 3000  
aaatcaaaat ggatgaaagg cttaaatcta aaacctcaaa ctttgcaact actaaaagaa 3060  
50 aacaccggag aaactctcca ggacattgga gtgggcaaag acttcttgag taattccctg 3120  
caggcacagg caaccaaaagc aaaaacagac aaatgggac atatcaagtt aaaaagcttc 3180  
tgccagcaa aggaacaat caacaaagag aagagacaac ccacagaatg ggagaatata 3240  
tttgcaaaact attcatctaa caaggaatta ataaccagta tatataagga gctcaaaacta 3300  
ctctataaga aaaaacacta ataaagctgat tttcaaaaat aagcaaaaaga tctgggtaga 3360  
cattttctcaa aataagtcac acaaatggca aacaggcatc tgaatatgtg ctcaacacca 3420  
55 ctgatcatca gagaatgca aatcaaaact actatgagag atcatctcat ccagttaaa 3480  
atggctttta ttcaaaagac aggcataaac aaatgccagt gaggatgtgg ataaaaggaa 3540  
acccttgagc actgttgggt ggaatggaaa ttgctaccac tatggagaac agtttgaaag 3600  
ttctcaaaa aactaaaaat aaagctacca tacagcaatc ccattgctag gtatatactc 3660  
caaaaaaggg aatcagtgta tcaacaagct atctccactc ccacatttac tgcagcactg 3720  
60 ttcatagca ccaaggtttg gaagcaacct cagtgtccat caacagacga atggaaaaag 3780  
aaaaatgtgt gcacatacac aatggagtag tacgcagcca taaaaaagaa tgagatcctg 3840  
tcagttgcaa cagcatgggg ggcactgggc agtatgttaa gtgaaataag ccaggcacag 3900  
aaagacaaac ttttcatgtt ctcccttact tgtgggagca aaaaataaaa caattgacat 3960  
agaaaatagag gagaatgggt gttctagagg ggtgggggac aggttgacta gactcaaca 4020  
65 taattttatg tatgttttaa aataactaaa agagtataat tgggttgttt gtaacacaaa 4080

5 gaaaggataa atgcttgaag gtgacagata ccccatcttac cctgatgtga ttattacaca 4140  
 ttgtatgcct gtatcaaaat atctcatgta tgctatagat ataaacccta ctatattaaa 4200  
 aattaaaaat ttaattggcca ggcacggtgg ctcatgtccg taatcccagc actttgggag 4260  
 gccgaggcgg gtggatcacc tgagggtcagg agtttgaaac cagtctggcc accatgatga 4320  
 aacctctgtc ctactaaaaga tacaaaaaatt agccaggcgt ggtggcacat acctgtagtc 4380  
 ccaactactc aggaggctga gacaggagaa ttgcttgaac ctgggaggcg gaggttgagc 4440  
 tgagccgaga tcatgccact gcactgcagc ctgggtgaca gagcaagact ccatctcaaa 4500  
 acaaaaacaa aaaaaagaag attaaaaatg taatttttat gtaccgtata aatataact 4560  
 ctactatatt agaagttaaa aattaaaaaca attataaaag gtaattaac acttaacta 4620  
 10 aaataagaac aatgtatgtg gggtttctag cttctgaaga agtaaaagt atggccacga 4680  
 tggcagaaat gtgaggaggg aacagtggaa gttactgttg ttagacgctc atactctctg 4740  
 taagtgactt aattttaacc aaagacagggc tgggagaagt taaagaggca ttctataagc 4800  
 cctaaaacaa ctgctaataa tgggtgaaagg taatctctat taattaccaa taattacaga 4860  
 tatctctaaa atcgagctgc agaattggca cgtctgatca caccgtcctc tcattcacgg 4920  
 15 tgcttttttt cttgtgtgct tggagatttt cgattgtgtg ttcgtgtttg gttaaactta 4980  
 atctgtatga atcctgaaac gaaaaatggt ggtgatttcc tccagaagaa tttagatacc 5040  
 tggcaggaag cagggtggctc tgtggacctg agccaattca atcttcaagg gtctctggcc 5100  
 aagaccagg tgcaaggcag aggcctgatg acccgaggac aggaaagctc ggtatgggag 5160  
 gggcgatgag aagcctgcct cggtggtgag cagcgcatga agtgccctta ttacgcttt 5220  
 20 gcaagattg ctctggatgc catctggaaa agcgcgccag cgggaatgca aggagtcaga 5280  
 agcctcctgc tcaaacccag gccagcagct atggcgccca cccggcgctg tgccagaggg 5340  
 agaggagtc aagcacctcg aagtatggct taaatctttt ttccacctga agcagtgacc 5400  
 aaggtgtatt ctgagggaag cttgagttag gtgccttctt taaacagaa agtcattgaa 5460  
 gcacccctct caagggaaaa ccagacgccc gctctgcggt catttacctc ttctctctct 5520  
 25 ccctctcttg ccctcgcggt tctgatcgg gacagagtga ccccgctgga gcttctccga 5580  
 gcccgctgct aggacctctc tgcaaaagggc tccacagacc cccgccttg agagaggagt 5640  
 ctgagcctgg cttaataaca aactgggatg tggctggggg cggacagcga cggcgggagt 5700  
 caaagactta attccatgag taaattcaac ctttccacat ccgaatggat ttggatttta 5760  
 30 tcttaatat ttcttaaat tcatcaata acattcagga ctgcagaaat ccaagggcgt 5820  
 aaaaacaggaa ctgagctatg ttgccaagg tccaaggact taataacat gttcagaggg 5880  
 atttttcgcc ctaagtactt tttattggtt ttcataagggt ggcttagggg gcaaggga 5940  
 gtacacagg agaggcctgg gcggcaggc tatgagcacg gcagggccac cggggagaga 6000  
 gtccccggcc tggggaggctg acagcaggac cactgaccgt cctccctggg agctgccaca 6060  
 35 ttgggcaacg cgaaggcggc cacgctcgt gtgactcagg accccatacc ggcttctctg 6120  
 gccccccac actaaccag gaagtcacgg agctctgaac cgtgggaaac gaacatgacc 6180  
 cttgccctgc tgcctccctg ggtgggtcaa gggtaatgaa ggtggtgca ggaatggcc 6240  
 atgtaaatca cagactctg ctgatgggga cgttctctc catcattatt catctcacc 6300  
 cccaaggact gaattatcc agcaactctc tgggtgtga caagccatga caaaactcag 6360  
 40 tacaacacc actcttttac taggcacaca gagcacggsc cacaccctg atatatag 6420  
 agtccaggag agatgaggct gctttcagcc accaggctgg ggtgacaaca gcggctgaac 6480  
 agtctgttcc tctagactag tagacctcg caggcactcc cccagattct agggcctgg 6540  
 tgctgcttcc cgaggcgccc atctgcccct gagactcagc ctgggtgccc acactgaggc 6600  
 cagccctgtc tccacacct cgccctcag gctcagctt ctccagcagc ttcttaaac 6660  
 45 ctgggtgggc cgtgttccag cgctactgtc tcacctgtc cactgtgtct tgtctcagc 6720  
 acgtagctcg caggttctct cctcacatgg ggtgtctgtc tcttcccca acactcacat 6780  
 gcgttgagg gaggagattc tgcgcctccc agactggctc ctctgagcct gaaactggct 6840  
 cgtggcccc gatgcagggt cctggcgctc ggctgcagc tgacctccat tccagggcgc 6900  
 tccccgtctc ctgtcatctg ccggggcctg ccggtgtgtt cttctgtttc tgtgtcct 6960  
 50 tccacgtcca gctgcgtgtg tctctgccc ctagggctctc ggggttttta taggcatagg 7020  
 acggggcggt ggtgggccag ggcgctcttg ggaaatgcaa catttggtg tgaaagttag 7080  
 agtgctgtc ctccactagg tccacgggca caggcctggg gatggagccc ccgccaggga 7140  
 ccgcctctc tctgcccagc actttcctgc cccctccct ctggaacaca gagggtcagt 7200  
 ttccacaagc actaagcctc ctcttcccaa aagaccagc attggcacc ctggacattt 7260  
 55 gccccacagc cctgggaatt cacgtgacta cgcacatcat gtacacactc ccgtccacga 7320  
 ccgaccccc ctgttttatt ttaatagcta caaagcaggg aatccctgc taaatgtcc 7380  
 ttaacaacac tggttaaaca aacgggtcca tccgcacggg ggacagtcc tcacagtga 7440  
 gaggaacatg ccgtttataa agcctgcagg catctcaagg gaattacgct gactcaaac 7500  
 tgccacctcc atgggatacg tacgcaacat gctcaaaaag aaagaatttc accccatggc 7560  
 60 aggggagtggt ttaggggggt taaggacggt gggggcgga gctgggggt actgcacgca 7620  
 ccttttacta aagccagttt cctggttctg atggtattgg ctcagtattg ggagactaac 7680  
 cataggggag tggggatggg ggaaccggga ggctgtgcca cctttgccat gccgagtg 7740  
 cctgggcagg ataattgctc agagatgccc acgtcctgat tcccccaac ctgtggacag 7800  
 aacccgcccc gccccagggc ctttgaggt gtgatctccg tgaggacct gaggtctggg 7860  
 65 atccttcggg actacattga ggcgggaaa gtaatccagg ggttctggga agaggcgggc 7920  
 aggagggtca gaggggggca gcctcaggac gatggaggca gtcagtctga ggctgaaaag 7980

5 ggagggaggg cctcgagccc aggcctgcaa gcgcctccag aagctggaaa aagcggggaa 8040  
gggaccctcc acggagcctg cagcaggaag gcacggctgg cccttagccc accagggccc 8100  
atcgtggacc tccggcctcc gtgccatagg agggcactcg cgtcgccctt ctgcatgaa 8160  
gtgtgtgggg atttgagaa gcaacaggaa acccatgcac tgtgaatcta ggattatttc 8220  
5 aaaacaaagg tttacagaaa catccaagga cagggctgaa gtgcctccgg gcaagggcag 8280  
ggcaggcacg agtgatttta ttagctatt ttattttatt tacttacttt ctgagacaga 8340  
gttatgctct tgttgcccag gctggagtgc agcggcatga tcttggtcca ctgcaacctc 8400  
cgtctcctgg gttcaagcaa ttctcgtgcc tcagcctccc aagtagctgg gatttcaggc 8460  
gtgcaccacc acaccgcgtt aattttgtat ttttagtaga gatgggcttt caccatgttg 8520  
10 gtcagactga tctcaaatc ctgacctcag gtgatccgcc caccctcagcc tcccaagtg 8580  
ctgggattac aggcattgag cactgcacct ggctattta accatttta aacttccctg 8640  
ggctcaagtc acaccactg gtaaggagt catggagttc aatttcccc ttactcagga 8700  
gttaccctcc tttgatatt tctgtaattc ttcgtagact ggggatacac cgtctcttga 8760  
cataattaca gtttctgtga ccacctgtta tccccggga cccactgcag gggcagctgg 8820  
15 gaggctgcag gcttcaggtc ccagtggggg tgcctatctc cagtagaaac ctgatgtaga 8880  
atcagggcgc aagtgtggac actgtcctga atctcaatgt ctcagtgtgt gctgaaacat 8940  
ctagaaatta aagtcctacc ctccactct actgggattg agcccttcc ctatcccccc 9000  
ccaggggagc agggagttct ctactctctg tggaggagg aatgatactt tgttattttt 9060  
cactgctggt actgaatcca ctgtttcatt tgttggtttg tttgttttgt tttgagaggc 9120  
20 ggtttcactc ttgttgctca ggctggaggg agtgcaatgg cgcgactctg gcttacttga 9180  
gcctctgctc cccaggttca agtgattctc ctgcttccgc ctcccatttg cctgggatta 9240  
caggcacccc ccaccatgcc cagctaattt tttgtatttt tagtagagac ggggggtggg 9300  
ggggttcacc atgttgcca ggctggtctc gaacttctga cctcagatga tccacctgcc 9360  
tctgctcctc aaagtgtgg gattacaggt gtgagccacc atgcccagct cagaatttac 9420  
25 tctgtttaga aacatctggg tctgagtag gaagctcacc ccactcaagt gttgtggtgt 9480  
tttaagccaa tgatagaatt tttttattgt tgttagaaca ctcttgatgt tttactactgt 9540  
gatgactaag acatcatcag cttttcaaa acacactaac tgcacctata atactgggtg 9600  
gtcttctggg tatcagcaat cttcattgaa tgcgggagg cgtttctctg ccactgcacat 9660  
gggtttaatt actccagcat aatcttctgc tctcatttct tctcttccct cttttaaaat 9720  
30 tgtgttttct atgttggtct ctctgcagag aaccagtgtg agctacaact taacttttgt 9780  
tcgaaacaaat tttccaaacc gcccctttgc cctagtggca gagacaattc acaaacacag 9840  
ccctttaaaa aggccttagg atcactaagg ggatttctag aagagcgacc tgttaactcta 9900  
agtatattaca agacgaggct aacctccagc gagcgtgaca gcccaggagg ggtgcgaggc 9960  
35 ctgttcaaat gctagctcca taaataaagg aatttccctc ggagatttct gaaagtagga 10020  
aagggttaca ttaagggtgc gtttggttagc atttcagtgt ttgccgacct cagctacagc 10080  
atccctgcaa ggcctcgga gacccagaag tttctcgccc ccttagatcc aaacttgagc 10140  
aaccggaggt ctggattcct ggggaagtcct cagctgtcct gcggtgtgac cggggcccca 10200  
ggctctggagg ggaccagtg ccgtgtggtt tctactgctg ggctggaagt cgggctcctc 10260  
40 agctctcgag tccgaggctt ggagccaggt gcctggaccc cgaggctgcc ctccacctg 10320  
tcggggcggg atgtgaccag atgttgccct catctgccag acagagtgcc gggggccagg 10380  
gtcaaggccg tttgtgctgg tgtgagcgc ccggtgccc gcccagcaga gcgctcggtc 10440  
ccatttccca ccttttctcg acgggaccgc cccggtgggt gattaacaga tttggggtgg 10500  
tttgcctatg gtggggaccc ctgcgcctc gagaacctgc aaagagaaat gacgggctc 10560  
tgtcaaggag cccaagtcgc ggggaagtgt tgcaggagg cactccggga ggtcccgctg 10620  
45 gcccgtccag ggagcaatgc gtccctcggt tgcgtcccag ccgctctac gcgctcctg 10680  
cctcccttcc acgtccggca ttcgtggtgc ccggagcccg acgccccgcg tccggacctg 10740  
gaggcagccc tgggtctccg gatcaggcca gcggccaaag ggtcgccgca cgcacctgtt 10800  
cccagggcct ccacatcatg gcccctccct cgggttacc caccagctag gccgattcga 10860  
cctctctcgg ctggggccct cgctggcgct cctgcacctt gggagcgcca gcggcgcgcc 10920  
50 ggcggggaag cgcggccca acccccggtt ccgcccggag cagctgcgct gtcggggcca 10980  
ggccgggctc ccagtggatt cgcgggcaca gacgcccagg acccgctcc ccactgggct 11040  
gagggactgg ggacccgggc acccgctcct ccccttcacc ttcagctcc gcctcctccg 11100  
gcgggacccc gcccgtccc gacccctccc ggggtccccg cccagcccc tccgggccc 11160  
ccagccccct ccccttccct tccgcgcccc cgcctctccc tcgaggcgcg agtttcaggc 11220  
55 agcgtgctg cctgctgctc acgtgggaag ccctggcccc ggccaccccc gcgatg 11276

<210> 4  
<211> 104  
<212> DNA  
60 <213> Homo sapiens

<400> 4  
gtgggctccc cgggggtcgg cgtccggctg ggggtgaggg cggccggggg gaaccagcga 60  
catgaggaga gcagcgacag cgactcaggg cgcttcccc gcag 104

65



<400> 5  
 gttagggaggt ggtggccgtc gaggggccag gccccagagc tgaatgcagt agggggctcag 60  
 aaaaaggggagc aggcagagcc ctgtctctcc tgtctccatc gtcacgtggg cacacgtggc 120  
 ttttcgctca ggacgtcgag tggacacggt gatctctgcc tctgctctcc ctctctccca 180  
 gtttgcataa accttacggag ttccacttca cgttttgatg gacacgcggg ttccaggcgc 240  
 cgagctgaag cagctgaaca gaggaggtcg ggcgcggcag ttggagccggg ttccgcagca 300  
 tgggggagaag tgtctggaag cacagacgct ctggcgaggg tgctgcagg ttacctataa 360  
 tctctcttcg aatttcaagg gtgggaaatga gagggtggga cgagaacccc ctctctctcg 420  
 ggggtgggag taagggtttt gcagggtgac gtggtcagcc aatatgcagg ttgtgtttta 480  
 agattttaact gtgtgttgac ggcagggtcg gtggctcac ccgggtaatc ccagcacttt 540  
 gggaagctga ggcaggtgga tcacctgagg tcaggagttt gagaccagcc tgaccaacat 600  
 ggtgaaacccc tatctgtact aaaaatacaa aaattagctg ggcattgtgg ttgtgtcctg 660  
 taactccagc tactttggag gctgaggcag gagaatcact tgaacccgag aggcggaggc 720  
 tgcagtgagc tgagatttgt ccatgtgact accgctggg cgacaagagt gaaactctgt 780  
 ctttaaaaaa aaaaagtgtt cgttgattgt gccaggacag ggtagaggga gggagataag 840  
 actgtctctc agcacagatc ctgtgtccat ctcttaggtat gaagaggggc acatggggag 900  
 agagagagcc agatggctcc accctgtgag gaagggacag gtttctgagg gtgtcagggg 960  
 atgggtgctg tggggcctgc cgtgtcccca cctgttttt ctggatttga tgttgaggaa 1020  
 cctccgctcc agccccctt tggctccagc tgcctccagg ccctaccgtg ccgctagaa 1080  
 gaagctccga tttcaccctt tccccacaaa tccccaagac atgtaagact tccggccatg 1140  
 gacagaagga gggtagacct ctggggcgct tttttttct ttttttctt ttatgtgtgc 1200  
 aaaagtcata taacatgaga ttggcactcc taacaccggt ttctgtgtac agtgcagaa 1260  
 tgcataactc gcggtgttta cagcaggttg cttgaaatgc tgcctcttgc gtgactggaa 1320  
 gtccctaccc atcgaaaggc agctgcctca cactctgtgc ggcctcaggt gaccacggcg 1380  
 agtctagata ccgtcatgca acccagtttt gctttttgtg ctccagcttc ctctgttag 1440  
 gagagtttga gttctctgat caggactctg cctgtcattg ctgttctctg acttcagatg 1500  
 aggtcacaaat ctgccccctg cttatgcagg gagttagggc tgggtccccg gtgtccccgt 1560  
 cagctgcagg gtgagtggg gttgtccccc aggtgtccct gtcacgtgta gggtagagta 1620  
 ggcgcgggcc ccgggtgtcc ctgtcccctg cagcgtgatt gagggtgtgc ccccggggtg 1680  
 cctctgtcac gttagggtag gtgaggcgcc atccccgggt gtctctgtca cgtgtagggt 1740  
 gagtgcaggc tgggtccccg ggtgtccctg ccggtgcagg gtgagtgcag caactgtccc 1800  
 ggggtgtccc gtcatgtga ggttagagta ggcgcggctc ccgggtgtcc ctctcagggt 1860  
 tagggtagt gaggcgcgcc cccagggtgt cctctgtcac gttagggtag gtgaggcacc 1920  
 gtccccgggt gtccccctga ggtatagggt gagttaggca ctgtccccg gtgtccccgt 1980  
 cagctgcagg gtgagtgcag ccggcccccc ggtgtgtccc cttaggtgca gggtagagta 2040  
 ggcgcgtgcc ctgggtgtcc ctgtctctgt tagggtagt gaggtctgt cccagggtgt 2100  
 ccttggcggt ttgtccactg agcttgtctc tgaattgtt ctcttctcat agccacagct 2160  
 gcgcgggtgt cccattgcct gggtagatgg tgcaggcgca gtgtgtgtcc caagcccat 2220  
 ctttctgat gctcgtctct tcttggtaac ctctccgttc catttgcata cggggacacg 2280  
 ggactgcagg ctctcgcttc ccgcgtgcca ggcactgcag ccacagcttc aggtccgctt 2340  
 gccctctgtt ggcctgggct gctaccacgc tgccccccac atgcatgtgt ccaatactcc 2400  
 tctccagagt tgcctatcgc cgaagctgga ctctgggctg ccgtgtctgt ctgcacagt 2460  
 ttgttgaga catcccagaa agggttctct gtgccctgaa ggaagcaag tcacccagc 2520  
 cccctcagct gtctctgttt ctcccaggct gcccccttgc ttggccccct tgggtgggtg 2580  
 gcaacgcttg tcaccttatt ctgggcacct gcgcctcatt gcttaggctt ggctctgctc 2640  
 ccagtccccc cctcacatgg attgacgtcc agccacaggt ttgaggtgtc ctgtctgtct 2700  
 cctgtcttga gacccacgtg gagggcgggt gtctccgcca gcttctgtca gacttccctc 2760  
 ttgggtctta gttttgaatt tcactgattt acctctgagc ttcttatctc tccattgtat 2820  
 gctttttctt ggtttattct tttactctt ttctagcttc ttagtttagt catgccttcc 2880  
 cctctaaagt ctgccttacc tgcaccctgt gttttgatgt gaagtaatt caacatcagc 2940  
 cactttcaag tgttctttaa atacttcaaa gtgttaatac ttcttttaag tattcttatt 3000  
 ctgtgatttt tttcttttgg cacgctgtgt ttttacgtga aatcatcttg atatcagta 3060  
 ctttaagta tttcttagct tattctgtga tttcttgag cagttagtta ttgcaacct 3120  
 gtttatgttc aagatatgta gagtatcaag atacgtagag tatttttaag tatcatctta 3180  
 ttatgtatt ctaacctcagt tgtgtagtgc tctgtataat accaattatt tgaagtttgc 3240  
 ggagccttgc tttgtgatct agtgtgtgca tgggttccag aactgtccat tgtaaaattg 3300  
 acatccttgc aatagttggc atgcattgtc acctatatca gcttattaag gtccagtgca 3360  
 aagcttctgt ctccctctag atcatgaaa ttccaagaa gaggccatag tccctcaact 3420  
 gggggatggg tctgttcaat tcttctctgt tggtagcatt ttgtgaggc attgttaagt 3480  
 qcatgcacgt ggtagaattt tcttctctcc gatcagtgaa tcttttggag acttctatgt 3540

ctctagtaat ctagtaaatc tttttttaa ttgctcttag tactgccaca ctgggcttct 3600  
 tttgattagt attttctatc tgtgtctgtt ttctgccttt aatttatata tatatatata 3660  
 tttttttttt ttttgagaca gagtcttggc ctgtcgccca ggggtgagtc agtgggtgta 3720  
 5 tcacaggcca gtgtaacttt taccttctgg cctgagccgt cctctcacct cagcctccctg 3780  
 agtagctgga actgcagaca cgcaccgcta cactcggtta atttttaaat tttttctgga 3840  
 gacagggctc tgctgtgttg ccacggctgg tctcaaactc ttggactcaa gggatccatc 3900  
 tacctcggct tcccaaagtg ctgaattaca ggcagtagcc accatgtctg gcctaatttt 3960  
 caacactttt atattcttat agtgtgggta tgcctgttta acagcatgta ggtgaatttc 4020  
 10 caatccagtc tgacagtcgt tgtttaactg gataacctga tttattttca tttttttgtc 4080  
 actagagacc cgcctgggtc actctgattc tccacttgcc tgttgcatgt cctcgttccc 4140  
 ttgtttctca ccacctcttg ggttgccatg tgcgtttcct gccgagtggt tgttgatcct 4200  
 ctctgttcct cctgggtcact gggcatttgc ttttatttct ctttgcttag tgttaccctc 4260  
 tgatcttttt attgtcgttg tttgcttttg tttattgaga cagtctcact ctgtcaccca 4320  
 ggcctggagt taatggacaca atctcggtc actgcaacct ctgcctcctc ggttcacaga 4380  
 15 gtctctcatt ctcaacctca tgagtagctg ggattacagg cggccaccac cagcctggc 4440  
 taatttttgt atttttagta gagataggct ttaccatgt tggccaggct ggtctcaaac 4500  
 tctgacctc aagtgatctg cccgccttgg cctccacag tgcctgggtt acaggtgcaa 4560  
 gccacctgac ccggcatacc ttgatctttt aaaatgaagt ctgaaacatt gctacccttg 4620  
 20 tccctgagcaa taagacctct agtgtatttt agctctggcc acccccagc ctgtgtgtctg 4680  
 ttttccctgc tgacttagtt ctatctcagg catcttgaca ccccacaag ctaagcatta 4740  
 ttaatatatt tttcgtgtt gagtgtttct gtatgtttgc ccccgctctc ctttccctcc 4800  
 tttgttcccc gtctgtcttc tgtctcaggc ccgctgtctg ggggtccctt ccttgtcctt 4860  
 tgcgtgtgtc ttctgtcttg ttattgtctg taacccctag ctttacctgt gctggcctcc 4920  
 atggcatcta gcgacgtccg gggacctctg cttatgatgc acagatgaag atgtggagac 4980  
 25 tcacggaggg ggcgtgtcat ttggccctg agtgtctgga gcaccacgtg gccagcgttc 5040  
 cttagccagt gagtgcacgc aacgtccgct cgcctcggtt tcagcctgga aaacccagc 5100  
 catgtcgggg tctgtgtggt ccgctgtgtc gaggttgaaa tcgcgcaaac ctgctgtgtg 5160  
 gcgcccagctc tgacgggtgt cctggcggtg gagggtgtct cttcctccct tctgtctggg 5220  
 aaccaggaca aaggatgagg ctccgagccg ttgtcgccca acaggagcat gacgtgagcc 5280  
 30 atgtggataa ttttaaaatt tctaggctgg gcgctgtggc tcacgcctgt aatcccagca 5340  
 ctttgggagg ccaaggcggg tggatcacga ggtcaggagg tcgagaccat cctggccaa 5400  
 atgatgaac cccatctgta ctaaaaacac aaaaattagc tgggctgtgt ggcgggtgcc 5460  
 tgtaatccca gctactcggg aggttgaggc aggagaattg cttgaacctg ggagtggaa 5520  
 gttgcagtga gccgacattg caccactgca ctccagcctg gcaacacagc gagactctgt 5580  
 35 ctcaaaaaaa aaaaaaaaaa aaaaaaaaaa aattctagta gccacattaa aaaagttaaa 5640  
 aagaaaaggt gaaattaatg taataataga ttttactgaa gccagcatg tccacacctc 5700  
 atcatttttag ggtgttatag gtgggagcat cactcacagg acatttgaca ttttttgagc 5760  
 tttgtctgct ggatccctgt tgtaggctcc gtgcgtggcc atctcggcct ggacctgctg 5820  
 40 ggcctcccat ggccatggct gttgtaccag atggtgcagg tccgggatga ggtcgccagg 5880  
 ccctcagtga gctggatgtg cagtgtcccg atggtgcacg tctgggatga ggtcgccagg 5940  
 ccctgctgtg agctggatgt gtggtgtctg gatggtgcag gtcagggtgt aggtctccag 6000  
 gccctcgggt agctggaggc atggagtcag gatgatgcag gtccgggtgt aggtcgccag 6060  
 gccctgctgt gagctggatg tgtggtgtct ggatgggtga ggtcagggtgt gaggtctcca 6120  
 45 ggcctcgggt aagctggagg tatggagtcc ggatgatgca ggtccgggtgt gaggtcgcca 6180  
 ggcctctgtg tgagctggat gtgtggtgtc tggatggtgc aggtctgggt tgaggtcacc 6240  
 aggcctctgc gtgagctggg tgtgcggtgt ctggatggtg cagggtctga gtaggtctgc 6300  
 cagacggtgc cagaccatgc ggtgagctgg atatgcgggt tccggatggt gcaggtctg 6360  
 ggtgaggttg ccaggccctg ctgtgagtgt gatgtgggtt gtcgggatgc tgcaggctcc 6420  
 50 gtgtgaggtc accaggccct gctgtgagct ggatgtgtgt tgtctggatg gtgcaggctc 6480  
 ggggtgaagg tcgcccaggcc cctgcttctg agctggatgt gttgtgtctg gatgtgtgag 6540  
 gtctggagtg aggtcgccag gccctcgggt agctggatgt gcagtgtcca gatgtgagc 6600  
 gtccgggtgt aggtcgccag accctcgggt gagctggatg tgcgtgtctt ggtggtgca 6660  
 ggtctggagt gaggtcgcca ggcctcgggt gagctggatg tatggagtcc ggtggtgtcc 6720  
 55 ggtccgggtg gaggtcgcca gacctgtctg tgagctggat gtgcgggtgt tggatggtac 6780  
 aggtctggag tgaggtcgcc agacctgtct gtgagctgga tatgcgggtgt ccggatggtg 6840  
 caggtcaggg gtgaggtctc caggccctcg gtgagctgga ggtatggagt ccggatgatg 6900  
 caggtccggg gtgaggtctc caggccctgc tgtgaactgg atgtgcggcg tctggatggt 6960  
 gcaggtctgg ggtgtggtcg ccaggccctc ggtgagctgg aggtatggag tccggatgat 7020  
 gcaggtccgg ggtgaggtcg ccaggccctc ctgtgagctg gatgtgcggc gctcggatg 7080  
 60 tgcaggtctg ggtgtgtgtc gccaggccct cggtgagctg gaggatgga gtcggatga 7140  
 tgcaggtccg ggtgtgaggt gccaggccct gctgtgagct ggtgtgtctg tatccggatg 7200  
 gtgcagtcgg ggtgtgaggt gccaggccct gctgtgagct ggtgtgtctg tatccggatg 7260  
 gtgcaggtct ggggtgaggt caccaggccc tgcggtagag tgggtgtgag gttccgggt 7320  
 65 gctgcaggtc cgggtgaggt tgcggagggc ctgcgtgagc tggatgtgag gttcccggt 7380  
 gtccggatgg tgcaggtcca ggtgtgaggt gctaggccct tgggtgggtg gatgtgctg 7440

gtccggatgg tgcaggtctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgagg 7500  
gtctgcatgg tgcaggtctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgagg 7560  
gtccggatgg tgcaggtccg gcgtgaggtc gccaggccctt gctgtgagct ggatgtgagg 7620  
tgtctggatg gtgcaggtcc ggggtgaggt agccaaggcc ttcggtagc tggatgtggg 7680  
5 gtgtccggat ggtgcaggtc cggggtaggg tgcagggcc ctgaggtag ctggatatgc 7740  
gggtgtagga tgggtgaggt cggggtaggg gtccaggcc cctgaggtag gctggatgtg 7800  
cgggtgtagg atggtgaggt tccgggtgga ggtgaggtc cctgagtag agctggatgt 7860  
gctgtatccg gatggtgag gtccgggtg aggtgaggt gccctgagct gagctggatg 7920  
tgctgtatcc ggatggtgca ggtctggcgt gaggtagcca ggccctgagg ttagctggat 7980  
10 atgagggtgc ggatggtgca ggtccgggtg gaggtagcca ggccctgagg ttagctggat 8040  
gtgcgggtgc cggatggtgc aggtctgggg taggtgagc aggcctgctg gtgagctgga 8100  
tgtgctgtat ccggatggtg caggtagggg gtgaggtgag caggccctgc ggtgagctgg 8160  
atgtgctgta tccggatggt gcaggtaggt cgtgaggtgc ccaggccctg cgtgagctgg 8220  
gatgtgagct gtacggatgg tgcaggtccg gggtagaggt gccaggccct gcggtagggc 8280  
15 gtatgtgtgt tgcctggatg gtgcaggtcc ggggtgaggt cggcaggccc tggtagagc 8340  
tggatgtgtg gtgtctggat gctgcaggtc cggggtaggt tgcagggcc ctgtagagc 8400  
tggatagctg gtgtcccggt gtccgaatgg tgcaggtcca gggtagaggt gccaggccct 8460  
tggtagggctg gatgtgaggt gtccggatgg tgcaggtcgt gggtagaggt gccaggccct 8520  
20 tggtagagct gatgtgagg gtccggatgg tgcaggtccg gggtagaggt accaggccct 8580  
cggtagatct gatgtggcat gtccctctcg ttttaag 8616

<210> 6  
<211> 2089  
<212> DNA  
25 <213> Homo sapiens

<400> 6  
gtactgtatc cccacgccag gcctctgctt ctccaagtcc tggaaaccca gcccgccctc 60  
agcatgccc tgtctccact tgcctgtgct tccctggctg tgcagctctg ggtggggagc 120  
30 cagggggccc gtccacaggc tgggtccaagt ggattctgtg caaggctctg actgcctgga 180  
gctcacgttc tcttacttgt aaaatcagga gtttgtgcca agtggctctt aggggtttga 240  
aagcagaagg gatttaaat agatggaaac actaccacta gcctccttgc cttcccttgg 300  
gatgtgggtc tgattctctc tctctttttt tttctttttt tgagatggag tctcactctg 360  
ttgcccaggc tggagtgagc tggcataatc ttggctcact gcaacctcca cctcctgggt 420  
35 ttaagcgatt caccagcctc agcctcctaa gtactggga ttacaggcac ctgccaccac 480  
gcctggctaa tttttgtact ttttagagag acgggggttc accatgttgg ccaggctggg 540  
ctcgaactca tgacctcagg tgatccacc accctggcct cccaaagtgc tgggtttaca 600  
ggctaagcca ccgtgcccag cccccgattc tcttttaatt catgtgttgc tgtatgaatc 660  
40 tccaactat tgatttagg tcatgagagg ataaaaatccc accacttgg cgactcaact 720  
cagggagcac ctgtgcagg agcacctggg gataggagag ttccaccatg agctaacttc 780  
taggtggctg catctgaatg gctgtgagat tttgtctgca atgttcggct gatgagagt 840  
tgagattgtg acagattcaa gctggatttg catcagtgag ggacgggagc gctggctctg 900  
gagatgccag cctggctgag cccaggccat ggtattagct tctccgtgct cggcccaggc 960  
tgactgtgga gggctttagt cagaagatca gggcttcccc agctccctg cactctgag 1020  
45 tccctggggg gccttgtgac acccatgccc ccaaatcagg atgtctgag agggagctgg 1080  
cagcagacct cgtcagaggt aacacagcct ctgggctggg gaccccgagc tggtagctgg 1140  
gccatttctc tgcattctgg ggagggtcag ggctttccct gtgggaacaa gttataacac 1200  
aatgcacctt acttagactt tacacgtatt taatgggtgt cgacccaaca tggctatttg 1260  
accagtattt tggaaagaat ttaattgggg tgaccggaag gagcagacag acgtgggtgt 1320  
50 ccccaagatg ctctctgtca ctactgggac tgtgtgtctg cctggggggc cttggaggcc 1380  
cctcctcccc ggacagggtg ccgtgccttt tctactctgc tgggctctgc gctgtaggct 1440  
agggcaccag ctccggagca cccgagggcc cagtgtccac ggagtggcag gctgtcagcc 1500  
acagatgccc aggtccaggt gtggccgctc cagccccctg gcccccagtg gttgttttgg 1560  
gggaaaaggc caagggcaga ggtgtcagga gactgggtgg ctcatgagag ctgattctgc 1620  
55 tcttggctg agctgacctg agcagcctct cccgccccct ccatctgaag ggtgtggct 1680  
ctttctacct gggggctcct cctggggcca gccttgggct accccagtgg ctgtaccaga 1740  
gggacaggca tctgtgtgtg aggggcatgg gttcacgtgg cccagatgc agcctgggac 1800  
caggctcccc ggtgctgatg gtgggacagt caccctgggg gttgaccgcc ggactggggc 1860  
tccccagggt tgactatagg accaggtgtc caggtagcct gcaagtagag gggctctcag 1920  
60 aggcgtctg ctggcatggg tggacgtggc cccgggcatg gccttcagcg tgtgctgccc 1980  
tgggtgacct gagccctcac tgaagctggg ggggctgtg gcttccccct agcttcccc 2040  
tagtctgttg tctggctgag caagcctcct gaggggctct ctattgagc 2089

<210> 7  
<211> 687  
<212> DNA  
<213> Homo sapiens

5

<400> 7  
gtggctgtgc tttggtttta ctctcttttt aaacagaagt gcgtttgagc cccacatttg 60  
gtatcagctt agatgaaggg cccggaggag gggccacggg acacagccag ggccatggca 120  
cggcgccaac ccatttgtgc gcacagttag gtggccgagg tgccggtgcc tccagaaaag 180  
10 cagcgtgggg gtgtaggggg agctcctggg gcagggacag gctctgagga ccacaagaag 240  
cagccggggc agggcctgga tgcagcacgg cccgagggtc tggatccgtg tccgtgctgtg 300  
gtgcgcagcc tccgtgctgt tccgtttacg gggcccgggg accagggcac gactgccagg 360  
agccacccgg gctctgagga tccctggacct tgccccacgg ctctgcacc ccacccctgt 420  
ggctgcgggt gctgcgggtg ccccgtcacg tgaggagagt gtggggtgag gtggacagag 480  
15 gtgtggcatg aggatcccggt gtgcaacaca catgcggcca ggaacccgtt tcaaacaggg 540  
tctgaggaag ctgggagggg ttctagggtc cgggtctggg tggctgggga cactggggag 600  
gggctgcttc tcccctgggt ccctatggtg ggggtgggac ttggccggat ccactttcct 660  
gactgtctcc catgctgtcc ccggcag 687

20

<210> 8  
<211> 494  
<212> DNA  
<213> Homo sapiens

25

<400> 8  
gtgggtgccg gggacccccg tgagcagccc tgctggacct tgggagtggc tgctgattg 60  
gcacctcatg ttgggtggag gagggtactcc tgggtgggccc gcagggagtgc caggtgaccc 120  
tgtcactgtt gaggacacac ctggcaccta ggggtggaggc cttcagcctt tccctgcagca 180  
catggggccc actgtgcacc ctgactgccc gggctcctat tcccaaggag ggtcccactg 240  
30 gattccagtt tccgtcagag aaggaaaccgc aacggctcag ccaccaggcc ccggtgcctt 300  
gcacccagct cctgagccag ggggtctcctg tccctgaggct cagagagggg acacagcccc 360  
ccctgccctt ggggtctgga gtgggtggggg tcagagagag agtggggggac accgccaggc 420  
cagggccctga gggcagaggt gatgtctgag ttctctgctg gccactgtca gtctcctcgc 480  
ctccactcac acag 494

35

<210> 9  
<211> 865  
<212> DNA  
<213> Homo sapiens

40

<400> 9  
gtaagggtca cgtgtgtag tctgttccag gatgtgtgtc tctgggatat gaatgtgtct 60  
agaatgcagt cgtgtctgtg atgcgtttct gtgggtggagg tacttccatg atttacacat 120  
ctgtgatatg cgtgtgtggc acgtgtgtgt cgtggtgcat gtatctgttg cgtgcatatt 180  
45 tgtgtgtgtg gtgtgtgtgg cacgtgtgtg tccatgggtg gtgtgcctgt ggtgtgcatg 240  
tgtgtgtgtc tgtgacacgt gcatgttcat gctgtgtgct gcatgtctgt gatgtgccta 300  
tttgtgtgtg gtgtgtgcat gtgtccgtga catatgcgtg tctatggcat ggggtgtgtg 360  
ggcccccttg ccttactcct tccctcctcca ggcattggtc gcaccattgt cctcacgctc 420  
tcgggtgctg gtttggggag ctccacattc agggctcctca cttctagcat ggggtgcccc 480  
50 gtcctgtcac agggctgggc cttggagact gtaagccagg tttgagagga gaggtagggat 540  
gctggtggtg ccttccctgga cccctggcac ccccaggacc ccagtctggc ctatgccggc 600  
tccatgagat ataggaaagg tgattcaggc ctgcgtcccc gggacacact cctcccagag 660  
cggccggggg ccttggggct cggcaggggt gaaagggggc ctgggcttgg gttcccaccc 720  
55 agtgggtcatg agcacgctgg aggggtgaag cctcaaagtc gtgccaggcc ggggtgcaga 780  
ggtaagaag tatccctgga gcttcggtct ggggagaggc acatgtggaa acccacaagg 840  
acctctttct ctgacttctt gagct 865

60

<210> 10  
<211> 3782  
<212> DNA  
<213> Homo sapiens

65

<400> 10  
tgtgggattg gttttcatgt gtgggatagg tggggatctg tgggattggt ttttatgagt 60  
ggggtaacac agagttcaag gcgagcttct tccctgtagt gggctctgcag gtgctccaac 120

agctttattg aggagacat atcttccctt gaactatggt cgggtttata gtaagtcagg 180  
 ggtgtggagg cctcccctgg gctcccctgt ctgtttcttc cactctgggg tegtgtgggt 240  
 cctgtgtggg tgtgtggcgg gtgggcaggg ctccaggcc tccttgtgtt cattggcctg 300  
 gatgtggccc tggctacgct cgtcccttgg aattcccctg cgagtggag gctttcttcc 360  
 5 tttctttttt tctttctttt tttttttttt tgataacaga gtctcgctct tttttgcca 420  
 ggtctggagt gtttggcgtg atcttggctc actgcaacct gtgtctcctg agttcaagca 480  
 attctcttgc ctacgcctcc caagtagctg gaattatagg cccccaccac catgctgact 540  
 aatttttgta attttagtag agacgagggt tctccatgtt gggcaggctg gtctcgaact 600  
 cctgacctca ggtgatcctc ccacctcggc ctcccaaagt gctgggatga cagggtgtaa 660  
 10 ccgcccgcgc cggccgagac tcgcttcctg cagcttcctg gagatctgca gcgatagctg 720  
 cctgcagcct tgggtgctgac aacctcctgt ttccttctcc aggtctcgtc aggggtctct 780  
 ccatttcatg actctcttca cagaagagtt tcacgtgtgc tgatttcccg gctgttctct 840  
 gcgtaatggt tgtctgctgt ttatcgatgg cctccttcca tttcttttag gctttgttta 900  
 ttgttgtttt tccggctcct tgaaggaaaa gtttcgatta tggatgtttg aacttctctt 960  
 15 tctaaacaag catctgaagt tgcggttttc cctctaaagc agggatcccg agggccctgg 1020  
 ctgtggagtg gcaccgctct ggggcctgtt aggaaccccg cgcacagcgg gaggctagggt 1080  
 ggggtgtggg gagccagcgt tcccgcttga gccccgccc tctcagatca gcagtggcat 1140  
 gcggtgctca gaggcgcaca caccctactg agaactgtgc gtgagagggg tctagatctc 1200  
 gtgtctctta tgggaatcta atgectgatg atctgaggtg gaaccgtttg ctcccaaac 1260  
 20 catccccctc cccactgctg tctgtggaa aaactgtctt ccacgaacc agtccctggg 1320  
 accacaatgg ttggggaccc tgtgttaaag acctgtctca gcagcctctc gtcagtgttg 1380  
 atattatggc tttctgtgtg tgagtccaga ataattacgg atttctgtga tgccttccgc 1440  
 cgacctcaga cccatgggct atttgtgggc gtgttgcctg ctccctgggtt gggaaagggtg 1500  
 caggcccat gtaccttctt gttactgcct tccagggttg. tcttcagggt tgaatcgta 1560  
 25 tccatgtggg tttagccac ggcctcctcg ccagctcctg ggggctgggg aacatgctga 1620  
 agcacagagt caccgtgcgc gtcttttgat gcttcacaag ctccaggcct cctgtgtccg 1680  
 tgttagtggt tgtcacgtgc ctgctcacat cctgtcttgg ggacgcaggg gcttagcagg 1740  
 tcccgtagta aatgacaagc gtccctgggg agtctgcaga ataggaggtg ggggtgcagg 1800  
 tctctctccc gctcttccag actcttctcc tgctgtgtc gtggctgcac ctgcactccc 1860  
 30 gcaatccctc cagcactggg ctggagaggc ccgggagctc gactgtccact tgtgccactg 1920  
 gactgtggat ggcagtcggt caggggggtc tgatgtgtgg tgactgtgga tggcggttgg 1980  
 tcacaggggt ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgact 2040  
 gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga 2100  
 tgtgtgact gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcgtgtctg 2160  
 35 ggggtctgat tgggtactgt ggtggcagt cgtggggtct gatgtgtgtg gactgtggat 2220  
 ggcgtgtctg ggggtctgat tgggtactgt ggtggcagt cgtggggtct gatgtgtgtg 2280  
 gactgtggat ggcgtgtctg ggggtctgat tgtgtgact gtggatggcg gtctgtgggt 2340  
 ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgtgtg ctgtggatgg 2400  
 40 cgggtctggt gtctgatgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgtgtg 2460  
 ctgtggatgg tgatcggtca caggggtctg atgtgtgtg actgtggatg gcggtctgtg 2520  
 ggtctgatgt gtgtgtactg tggatgtgtg tgggtcacag ggtctgtatg tgtgtgact 2580  
 gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgg atggcggtcg gtcccggggg 2640  
 tctgatgtgt ggtgactgtg gatggcgatc ggtcacaggg gtctgtgtgt tgggtactgt 2700  
 ggatggcggt cgtggggtct gatgtgtgtg gactgtggat ggcgtctgtg gggctctgat 2760  
 45 tgtgtgact gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcgtctgtg 2820  
 ggggtctgat tgggtactgt ggtggcggt cgtggggtct gatgtgtgtg gactgtggat 2880  
 ggcgtgtgtt cccgggggtc tgatgtgtgt tgactgtgga tggcggtcgt ggggtctgat 2940  
 gtgtgtactg tggatggcag tctgtgggtc tgatgtgtgt tgactgtgga tggcggtcgt 3000  
 50 gatggcggtc gtgggtgtgac tgtggatggc ggtcgtgggg tctgatgtgt ggtgactgtg 3060  
 ggtgactgtg gatggcggtc gtgggtctgt atgtgtgtgt actgtggatg gtgatcggtc 3120  
 acaggggtct gatgtgtgtg gactgtggat ggcgtctgtg ggggtctgat tgtgtgact 3240  
 gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcgtctgtg ggggtctgat 3300  
 tgtgtgact gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgt atggcagtcg 3360  
 55 gtccacagggt tctgatgtgt ggtgactgtg gatggcggtc gtgggtctgt atgtgtgtgt 3420  
 actgtggatg gcggtctgtg ggtctgatgt tggatggcg tctgtgggtc 3480  
 tgatgtgtgt tgactgtgga tggcggtcgt ggggtctgat gtgtgtactg tggatgggtg 3540  
 tccgtcacag ggtctgtatg tgtgttagct gcaggtggag tcccaggtgt gctgttagct 3600  
 60 actttgcgtc ctccggcccc cggccccctg tttcccaaca gaagcttccc aggcgtctct 3660  
 tgggttcat cccgccatcg ggttggccg caggtccaca cgtcctgac ggaagaaaca 3720  
 agtgcacagc tctggccggg gcagggcaca tttgtggctc atgcctctc ctctgccggc 3780  
 ag 3782

12 / 18

<210> 11  
<211> 980  
<212> DNA  
<213> Homo sapiens

5

<400> 11  
gtctggggcac tgccttgacag gggtggggcac ggactcccag cagtgggtcc tccctggggc 60  
aatcactggg ctcactgacc gacagactgt tggccctggg gggcagtggg gggaaatgagc 120  
tgtgatgggg gcatgatgag ctgtgtgcct tggcgaaatc tgagctgggc catgccaggc 180  
10 tgcgacagct gctgcattca ggcacctgct cactgttgac tgcgcgccct ctctccagtt 240  
ccgcagtgcc ttgtttcatg atttgctaaa tgtcttctct gccagttttg atcttgaggc 300  
caaaggaaag gtgtccccct cctttaggag ggcaggccat gtgtgagccg tgtcctgccc 360  
agctggcccc tcatgtctgg gtctgaggcc aaaggaaacg tgtccccctt cttaggaggc 420  
cgggccgtgt ttgagccacg ccccgctgag cgggcctctc agtgcctggg ctgtcccagc 480  
15 gggccctgtg ccttttgacg atgtgggtctg tccacgtggc cctgtggctc ttgacagatg 540  
cctgttagca ctgtctcgcc tctaggggac agtgcgtgct accgcatgag gctcagagac 600  
ctctggggca atttcttggt cctccagggt ggggggtggg gtggcctggg ctgctggggc 660  
ccagaccctg tggccggcag ctgggcagca actcctggat cacatatgcc atccggggcca 720  
cgggtgggctg tgtgggtgtg agcccagctg gaccacacag tggcccagag gagacgttct 780  
20 gtgtcacaca ctctgcctaa gcccatgtgt gtctgcagag actcggcccc gccagcccac 840  
gatggccctg cattccagcc cagccccgca cttcatcaca aacactgacc ccaaaaaggga 900  
cggagggtct tggccacgtg gtctctgctg tctcagcacc caccggctca ctcccatgtg 960  
tctcccgctc gctttcgacg 980

25

<210> 12  
<211> 2485  
<212> DNA  
<213> Homo sapiens

30

<400> 12  
gtgagtcagg tggccagggt ccattgcctt gcgggtggct gggcgggctg gcagggtctc 60  
tgctcacctc tctcctgccc ctccccact gnccttctgc ccggggccac cagagtctcc 120  
ttttctggcc ccgccccctt ccggctcctg ggtgacaggc tcccagggcc ccggaacatc 180  
ggctcggctt ggcgcagccg gagcgagca ggtgccacac gaggcctgga aatggcaagc 240  
35 ggggtgtgga gttgctcctg cgtggaggac gaggggcggg ggggtgtgtc gggtcagggt 300  
tgcgcccagc gtttgagcct gcagcttgtc agctccaagt tactactgac gctggacacc 360  
cggctctcac acgcttctat ctctctctcc cgatacaaaa ggattttatc cgattctcat 420  
tctctgctct gtctgtgtac ccccgcgagg gcgcgggctc ttctctctgt gactagattt 480  
cccatctgga aagtgcgggg ttgaccgtgt agtttgcctc tctcgggggg cctgtgggtg 540  
40 ccatggggca ggcggcctgg gagagctgcc gtcacacagc cactgggtga gccacactca 600  
cgggtgtaga gccacagtgc ctggtgccac atcacgtcct ctggatttta agtaaaacca 660  
cacacctccc ggcaggcctc tgcctgcgac cctgtgtgtg cctggggaga gtggtagcac 720  
ggaggaaatt cgtgcacact caaggtcatc agcaaggtca tccgcagtcg ggtggaacgt 780  
ggaggcctct ctctgggacg gtctccagcg gataaaggac tgtgcacagc ttcgggaagct 840  
45 tttatttaaa aatataacta ttaattattg cattataagt aatcactaat ggtatcagca 900  
attataatat ttattaaagt ataattagaa atattaagta gtacacacgt tctggaaaaa 960  
cacaaattgc acatggcagc agagtgaatt ttggccgagg gacacgtgtg cacatgtgtg 1020  
taagcggccc ccaggcccac agaattcgct gacaaagtca cctccccaga gaagccacca 1080  
cgggcctcct tctgtgtgtg gaattttatt aagatggatc aagtcacgta ccgtccacgt 1140  
50 gtggcagggc tttggggaaat gtgagggtgt gactgcgtcc tcatgcccctg acagacagga 1200  
ggtgactgtg tctgtcctgt ccttaggaca cggacaggcc cgaagctcta gtccccatcg 1260  
tggtccagtt tggcctctga ataaaaacgt cttcaaaacc tgttgcccca aaaactaaga 1320  
acagagagag tttcccatcc catgtgctca caggggcgta tctgcttgcg ttgactcgct 1380  
gggctggccg gactcctaga gttggtgctg gtgcttctgt gcaaaaagtg cagtccctct 1440  
55 gccatcact gtgatctctg caccagcaag gaaagcctct tttctttctt tctttttttt 1500  
ttttttgaga cggaacgtca ctgttgtctg cctgggcttg agtgcagtgg cgcgatctca 1560  
actcactgca acctccgctt cccgggttcc agcattttct ctgctcagc ctcccagca 1620  
gctgagatta caggcaccca cccctgcgc ctggctaat tttgtatttt tagtagagag 1680  
gggttttttc catgttgccc aggtgtgtct cgaactcctg acctcaggtg atccaccac 1740  
60 ctcggcctcc caaagtgtg ggattacagg tgtgagccat cacgccagc cggaaagcct 1800  
cttttaagg tgaccaccta tagcgttcc cgaataaac aggtcttgtt tttgcagtag 1860  
gctgcaagcg tctcttagca acaggagtgg cgtcctgtgg gctctgggga tggctgaggg 1920  
tcgctggca gccatgcctt ctgtgtgcac ctttaggttc caccgggcta tctgtctctc 1980  
actgtttgtc tgaaaacgca cccttggcat cctgttttgg agagtctctg ctctcgttg 2040  
65 gtcactgctg aactaggggc aaggttgtat ccgttggcgc gcagcggcta catgtagggt 2100

13 / 18

catgagtctt tcaccgtgga caaattcctt gaaaaaaaaa aaaggagtcc ggtaaagcat 2160  
 tcattccggg tcaagtgtct ggttctgtga ataaactcta agattttaaga aaccttaatg 2220  
 aaagaaaacc ttgatgattc agagcaagga tgtggtcaca cctgtggctg gatctgtttc 2280  
 agccgccccca gtgcatgggt agagtgggga gcagggaattg tttgttcaga ggtctcatct 2340  
 5 ggtatgtttc tgaggtgttt gccggctgaa tggtagacgt gtcgtttgtg tgtatgaggt 2400  
 tctgtgtctg tgtgtggctc ggtttgagtg tacgcatgtc cagcacatgc cctgcccgtc 2460  
 tctcacctgt gtcttccgcg cccag 2485

<210> 13  
 <211> 1984  
 <212> DNA  
 <213> Homo sapiens

<400> 13  
 15 gtgaggcctc ctcttcccca ggggggcttg ggtgggggtt gatttgcttt tgatgcattc 60  
 agtgttaata ttctctgtgc tctggagacc atgactgtc tgccttgagg aaccagacaa 120  
 ggttcagacc ccttcttggg atgaagccgc acgggagggg ttgcacagcc tgaggactgc 180  
 gggctccacg caggctctgt ccagcggcca tgtccagagg cctcagggct cagcaggcgg 240  
 20 gaggggcctc gccctgcagc atgagcatgt gaattcaaca ccgaggaagc acaccagctt 300  
 ctgtcacgtc acccagggtc cgttaggggc cttggggaga tggggctggt gcagcctgag 360  
 gccccacatc tcccagcagg cctcgcagag gtggcctgga ctgggcgcct cttcagccca 420  
 ttgcccctcc cacttgcatg ggggtctacac ccaaggacgc acacacctaa atatcgtgcc 480  
 aacctaatgt ggttcaactc agctggcttt tattgacagc agttactttt ttttttttaa 540  
 tactttaagt tctaggggtac atgtgcacga cgtgcagggt agttacatat gtatacatgt 600  
 25 gccatgttgg tgtgctgcac ccattaaact atcatttaca ttaggatatat ctctaatgc 660  
 tatccctccc cactcccccc atcccatgac aggccttggg gtgtgatgtt cccacccttg 720  
 tgtccaagtg ttctcattgt tcagtcccca cctgtgagtg agaacatgtg gtgtttgggt 780  
 ttcttttctt gcaatagttt gctcagagtg atggtttcca gcttcgtcca tgcctctaca 840  
 aaggacatga actcatcctt ttttatgact gcatagtatt ccgtgggtga tatgtgccac 900  
 30 attttcttaa tccagtctat catcgatgga catttgggtt ggttgcaagt ctttgcact 960  
 gtgaatagtg ccgcaataaa catacgtgtg catgtgtctt tatagcagca tgatttataa 1020  
 tcttttgggt atataccag taatgggatg gctgggtcaa atggattttc tagttctaga 1080  
 tcccttgagg atcaccacac tgtcttccac aatgggtgaa ctagtttaca cttccaccaa 1140  
 cagtgtaaaa gtgttctggt gctggagagg atgtggacag cagttatttt tttatgaaaa 1200  
 35 tagtatcact gaacaagcag acagtttagt aaggatgcgt caggaagcct gcaggccaca 1260  
 cagccaattc tctcgaagac tccgggtttt tccgtgcat cttttgaaac tctagctcca 1320  
 attatagcat gtacagtgga tcaaggttct tcttcattaa ggttcaagtt ctagattgaa 1380  
 ataagttat gtaacagaaa caaaaatttc ttgtacacac aacttgctct gggatttgga 1440  
 ggaaagtgtc ctcgagctgg cggcacactg gtcagccctc tgggacagga tacctctggc 1500  
 40 ccatggctat ggggagctgg gcttgggctt gagggtcaca cagtgcacca tggccagctt 1560  
 cctgtggata ggatctgggt ctccgatcat gctgaggacc acagctgccca tgcgtgtaaa 1620  
 gggcaccacg tggctcagag ggggagaggt tcccagcccc agctttctta cgtctctcag 1680  
 ttatttttcc ctaagagtct gagaagtggg gccgcgcttg atggccttcg tctgcttcca 1740  
 gctggcacag aattgcacaa gctgatggta aacctgagt acttataatg aatgaggaat 1800  
 45 tgcgtagca gttaactgta gagagctcgt ctgttggaag gaaatttaag tttttcattt 1860  
 aaccgctttg gagaatgtta ctttatttat ggctgtgtaa attgtttgac attcagctcc 1920  
 tcgtagacag atactacgta aaaagtgtaa agttaacctt gctgtgtatt ttcctctatt 1980  
 ttag 1984

<210> 14  
 <211> 1871  
 <212> DNA  
 <213> Homo sapiens

<400> 14  
 55 gtgaggcccc tgccgtgtgt ctgtggggac ctccacagcc tgtgggcttt gcagttgagc 60  
 cccccgtgtc ctgccccttg caccgcagcg ttgtctctgc caagtctctt ctctctgccg 120  
 gtgctggatc cgcaagagca gaggcgcttg gccgtgcacc caggcctggg ggcgcagggg 180  
 60 caccctcggg agggagtggt taccgtgcag gccctggctc tgcagagacg caccacaggt 240  
 acacacgtgg tgaatgcagg cgtgacctg gctcctgctg ctctttggaa agtcaagagt 300  
 ggcggctcct ggggccccag tgagaccccc aggagctgtg cacagggcct gcaggggcca 360  
 ggcggcagcc tccctcccag ggtgcacctg agcctgcgga gagcaggagc tgcgtgagtg 420  
 gctggccccac agcgttcgct gcggctcacgt tccctgcgtg ggttgtttgg gatcgggtgg 480  
 agaatttggg tttgctgagt gctgctgtct tgaaccacgg agatggctag gagggggttt 540  
 65 cagagttgat ttttgtgaat caaactaaaa tcaggcacag gggacctggc ctacgcacag 600

5 gggattgtcc aatgtggtcc ccctcaaggc cgccccacag agccgggtggg cttgttttaa 660  
 agtgcgattt gacgaggagc gagaacacct gaaagctgta aaggggaacc tcagaaaatg 720  
 tggccgccag ggggtggttc aggtgctttg ctgggctgtg tttgtgaaaa ccattttgga 780  
 cccgccctcc aagttccacc tccaggtcca cctccaggc ccgccctggg ctgggggtat 840  
 10 gcctggcggt ccttgtgccc cagcccgagc cacagcaggc tgtgcacatt taaatccact 900  
 aagattcact cggggggagc ccaggtccca agcaactgag ggctcaggag tcctgaggct 960  
 gctgagggga cagagcagac ggggaacgct gcttctgtgt ggcaagttcc tgagggtgct 1020  
 ggccaggagc gtggctcaga gtgtatgttg ggtcccacc gggggcagaa ctctgtctct 1080  
 gatgagtcgg cagccatgta acaggaaggc gtggccacag ggagctggga atgcaccagg 1140  
 15 ggagctgcgc agctggccga ggtcccaggc ccaggccaca ggaaggcagc ggggacgccc 1200  
 ggggccacag cagagggcgc aggaaggga ggggatgccc aggccagagc agaggctacc 1260  
 gggcacaggc gggctcccgc agctgggtga gcgaggctca tgactcggcg agggaaacct 1320  
 cttgacgtga agctgacgac tgggtgtgct cagctcacag cccagccagg tcccgcgctt 1380  
 gagcaggaac tcagaacctt cccctttgtc taaagcacag cagatgcctt caggggcatc 1440  
 20 agggaaaaac agggaaagtc gttgagaaac gtcttaaaag aaggtgggat ggtggcaatt 1500  
 tctgttccag attttagtct gcccgggacc acagatgagt ctataacggg attgtggtgt 1560  
 tgccatgggg acacatgaga tggaccatca cagagggcac tggggctgca cctcccatct 1620  
 gagtccctggc tgtcccgggt ccaggccagg tcttgcagc ctacactacc tgtccctgcc 1680  
 gggagacagg gaaagcacc cgaagtctgg agcagggctg ggtccaggct cctcagagct 1740  
 25 cctgccaggc ccagcacctt gctccaaatc accacttctc tgggggtttc caaagcattt 1800  
 aacaagggtg tcaggttacc tctgggtga cggcccgca tcctggggct gacattgccc 1860  
 ctctgcctta g 1871

<210> 15  
 <211> 3801  
 <212> DNA  
 <213> Homo sapiens

30 <400> 15  
 gtgagcgac ctggccggaa gtggagcctg tgcccgctg gggcaggtgc tgctgcaggc 60  
 ccgttgctgc caccctctgt tccgtgtggg gcaggcgact gccaatccca aagggctcaga 120  
 ggccacaggc tgccccctgt cccatctggg gctgagcaga aatgcattct tctgtgggag 180  
 tgaggggtgt cacaacggga gcagttttct gtgctatttt ggtaaaagga aatgggtcac 240  
 35 gacagcctggg tgacttgagg tgtcttcaga aagcagctct gatccgaacc caagacgccc 300  
 gggccctgct gggcgtgagt ctctcaaac cgaacacagg ggcctctgtg ggcagtgatc 360  
 cctctgaacc cagagacctg gggccctgct gggcgtgagt ctctccgaac ccagagactt 420  
 cagggccctt ttgggcgtga gtctctccgc tgtgagcccc acactccaag gctcatccac 480  
 agtctacagg atgcatgag ttcattgatca cgtgtgacct atcaggggac agggccatgg 540  
 40 tgtggggggg gtctctacaa aattctgggg tctgttttcc ccagagcccc agagctcaag 600  
 gccccctctc aggtctcagc acaaatgaat tgaagatgga cacagatgca gaaatctgtg 660  
 ctgtttcttt tatgaataaa aagtatcaac attccaggca gggcaagggt gctcacacct 720  
 ataataccag cactttggga gggcaggtg ggtggatcac ttgaggccag gagtttgagg 780  
 ccaacctaac caacatagt aaattccatt tctacttaaa aaatacaaaa attagcctgg 840  
 45 cctggtggca cagcctgtga gtccccgcta tgccggaggc tgaggcagga gaattcattt 900  
 aacccaggag gcagaggttg cagtgcagcc actgcactcc agcctgggca 960  
 acagagttag acttcatctt aaaaaaaaa aaaaaagtat cagattccca aaaccatagt 1020  
 ggacagggtt ttttttatct tgcctctcga taattttac tgggtgctgtg ctgaggccg 1080  
 gaaactgggg tgcttctctc tgaaggcag accttcatgg gaagagaaat aagtgggtga 1140  
 50 tgggtgttaa accagaggtt taaactgggg tccgtctgtt ctgagttaac agtccagatc 1200  
 tggactttgc ctctttccag aatgctccct ggggtttgct tcatggggga gcagcagggt 1260  
 tggacacctt cgtgatgggg gaggcagcag tgcagacgct ctcatgatgg gggagtggca 1320  
 ggtgcagaca ccttctgtga tgggtgccag catgtccctg ttgcagctcc ctccccacaa 1380  
 55 ggatgcccgt ctctctgtgt cccacagatc cctgcttccc tctcacagcc ttacctggct 1440  
 ctggcctcca ctggctttgt ctgcattgat tccacatttc ctgggctccc agcacctctt 1500  
 gcctctctcc aggcacctct gcagtgctgg ccataccagt cagctgtgaa ctgtccactg 1560  
 cttattttgc tccccatgaa atgtattttt taggacaggc acccttggtt ccagcctctg 1620  
 gcacagcatc agtgaatgtt attgaaggac aaaggacaga caaacaatc aggaaaatgg 1680  
 gttctctcta aacacattgc aaagccacag aggtcagtag aggatgggtg ggcacagggt 1740  
 60 catcagatgt ggggtccaat ccagaatatt ctgtgctccc aaaggccact tggtcagagt 1800  
 gtgtgcttgc agaggtggct ctaaaagctc agcagtgagg gcagtggttc gccatactca 1860  
 ggggtgaacct acatccctctg tgtctgaagt atacagcaga ggcttgaaag gcattctggg 1920  
 gaagaaaaa ggcaaaatga ttaagaaaag tgaagaaagga aaagtggtaa gatgggaatt 1980  
 65 tcttcttcca gatttttagt tcccaaacca cagctcagat ggtagaatgt ggtcagaact 2040  
 gatggacaga acaatagaac aaaacggaag ccctatctct cagaacctgt tgttaattgt 2100  
 gtatgtggca cagctgatgg aaaagagagt gtgtgtgtaa tttttttttc tgagaaaact 2160



5 gactggaagc aaataagttg tgtctttaca gcatatacca gagcagattc taggtagaag 2220  
 aggagacaca tgcaaacac accagcaaca gaaataaaac aaaagactca aaggggaagg 2280  
 aggtgaacgt tccctgggtt ggtgttgggg aaggacacac agggagggcg atgaaaccag 2340  
 tgaggcaacg ggcatgtgct tcaactgcaga gaaactcagc ttgcctgagc cacagtga 2400  
 atggccattc cctggagcgt ttgtgcacgt gatttattta aggcgcctcg tgaggtcctg 2460  
 cacattcatc ctctcacttt gttctcctaa ccacctgaga ggtagaggag gaaaggctcc 2520  
 agggggagcag ccgccttggg tcacccagct ggcaaaaggc atgcatgatt gcagcctggc 2580  
 ctctgtctcc ggggcccctg ctctgcccga ggacccca caagtccagc ccataggctc 2640  
 10 aggggtgagcc ggagcccag gtcgtgttgg ggatggctgt gaaagaagaa atggacgtct 2700  
 gatgcacact tgggaaggtc ctaccagcag cgtcaaagaa atgcatgtga aactgacagc 2760  
 gagacccatc cctcaaagaa acgcacgtga aactgatggc gagacctgtc cccatccctc 2820  
 atgtcggctc cttttctggg cttgccaaaga gccagcatca ggttgaggca agctgaaaag 2880  
 acttttctgg aaagcagctt gtttgcatgg aagtccctac aatgtcctgt gtcttccag 2940  
 taattccact tctgaagtga ccagacatta tcacgggtct tatttaccat ttccagtggt 3000  
 15 ccaggcaggg ggacttgcca cagcaagtca cgaacctgcc caaatacagg gctaaggaga 3060  
 tattatgcat cacaaaactt gctctgccat taacatttt tcaaagaatt ttgagaat 3120  
 gtttaattgg acaaaacgtt tatttcaatg tagcagtggt caaagctgga tgtaaaagaa 3180  
 cacaccccag gagcctgccc tgaatgtcat gtgtgttcat ctttgacat ggacatacat 3240  
 20 gggcagtgag tgggtgtgag gccctggagg acatcggtgg gatgcctcca tcctgcccct 3300  
 ctggagacac catgtgtgcc acgtgcactc actggagccc tgtttagctg gtgccacctg 3360  
 gctcttccat cctgagatt caaacacagt gagattcccc acgcccact cagtgttctc 3420  
 ccacaaaaaa cctgagtcac acctgtgttc actcgaggga cggccgggag ccagggtctc 3480  
 acagtttatt atgtgttttt ggctgagtta tgtgcagatc tcatcagggc agatgatgag 3540  
 25 tgcacaaaaa cggccgtgag aggtttggat aactcaaca tcaactagcca ggtcctgggt 3600  
 gagtttggtc atgcagagtc tggatggcat gtagcatttg gagtccatgg agtgagcacc 3660  
 cagccccctc gggctgcagc gcatgcccc ggcaggacaa ggaagcggga ggaaggcagg 3720  
 aggtctcttg gagcaagctt tgcaggaggg ggctgggtgt ggggcaggca cctgtgtctg 3780  
 acattcccc ctgtgtctca g 3801

30 <210> 16  
 <211> 880  
 <212> DNA  
 <213> Homo sapiens

35 <400> 16  
 gtgagcagc tgatggtcag cacagagttc agagtccagg aggtgtgtgc gcaagtatgt 60  
 gtgtgtgtgt gtgcgcgcgt gcctgcaagg ctgatgtgta ctggctgcac gtaagagtgc 120  
 acatgtacgc atatacacgt gagcacatac atgtgtgcat gtgtgtacat gaaggcatgg 180  
 40 cagtgtgtgc acaggtgtgc aagggcacaa gtgtgtgcac atgcgaatgc acacctgaca 240  
 tgcattgtgt ttctgtgcaca gtctgtgtgg cattcacgtg aggtgcatgc gtgtgggtgt 300  
 gcagtgtag tagcatgtgt gcacataaca tgtattgagg ggtcctcgtg ttccccccg 360  
 taggtcctca gcaccagtgc cactccttac aggatgagac ggggtcccaag gccttgggtg 420  
 gctgaggctc tgaagctgca gccctgaggg cattgtccca tctgggcatc cgcgtccact 480  
 cctctcctg tgggcttctg tgtccactcc cctctcctg tgggcattta catccactcc 540  
 45 actccctctc tctgtgtggc atcccgctcc actccccctc tctgtgggca tctgcttcca 600  
 cctccccctc ctgtgggcat ttgcgtccac tccctctcct ggttcccttc tgtcttggcc 660  
 gagcctcggg ggcaggcaga tgacacagag tcttgactcg cccagggtgg ttcgcagctg 720  
 ccgggtgagg gccaggcccg atttactctg gaagagggat agtttcttgt caaatgttcc 780  
 50 ctctttcttg ttccatctga atggatgata aagcaaaaag taaaaactta aaatcccaga 840  
 gagggttcta ccgtttctca ctcttctctg gcgactctag 880

55 <210> 17  
 <211> 3186  
 <212> DNA  
 <213> Homo sapiens

60 <400> 17  
 gtgagccgcc accaaggggt gcaggcccag cctccaggga ccctccgcgc tctgctcacc 60  
 tctgaccccg ggtttcacct tggaaactct ggggttttag ggcaaggaaat gtcttacgtt 120  
 ttccagtgtg ctgctgcctg tgacagttc tgttcgcgtg gctctgtgca aagcacctgt 180  
 tctccatctc tgggtagtgg taggagcccg tgtggcccca ggtgtcccca ctgtgcctgt 240  
 gcactggccg tgggagctca tggaggccat cccaggggcag caggggcatg gggtaaaag 300  
 atgtttatgg ggagctctta cagaggaggg tgggaagggt tctgaacagt agatgggaga 360  
 65 tcagatgccc ggaggatttg gggctctcagc aaagagggcc gaggtgggtg cagggtgagg 420  
 tcgctggccc cacccccggg aaggtgcagc agagctgtgg ctccccacac agcccgcca 480

gcacctgtgc tctgggcatg gctgtgctcc tggaaacgttc cctgtcctgg ctggtcaggg 540  
 ggtgccccctg ccaagaatcg acaactttat cacagaggga agggccaatc tgtggaggcc 600  
 acagggccag cttctgcctg gagttagggc aggtggtggc acaagcctcg gggctgtacc 660  
 aaagggcagt cgggcaccac agggccgggc ctcacctca acaggcctcc cgagccactg 720  
 5 ggagctgaat gccaggaggc cgaagccctc gcccctatgag ggctgagaag gagtgtgagc 780  
 atttgtgtta cccaggggcg aggtgcgcgc aattaccgtg cacacttgat gtgaaatgag 840  
 gtcgtcgtct atcgtggaac cccagcaagg gctcacggga gatttttcca tacaaggctc 900  
 gtaccatgaa aatgggtttt aaccagagtg cttgcgcctt catgctctgg caggaggaggc 960  
 10 agagccacag ctgcatgtta ccgcctttgc accagctcca gaggcttggg accaggctgt 1020  
 ctcagttcca ggggtgcgtc ggctcagacc gccctcctct ctgccttctc tctctgcctc 1080  
 aaatcttccc tcgtttgcat ctccttgacg cgtgcctggg ccctcgtgca agctgcttga 1140  
 ctccctttccg gaaacccttg ggggtgtgtg gatacaggtg ccaactgagga ctggagggtg 1200  
 ctgacactgt ggttgacccc aggggtccagc tggcgtgctt ggggcctcct tgggccaatga 1260  
 15 tgaggtcaga ggagttttcc caggtgaaaa ctcctgggaa actcccaggg ccatgtgacc 1320  
 tgccacctgc tcctcccata ttcagctcag tcttgcctc atttcccac cagggtctct 1380  
 agctccgagg agctcccgtg gagggcctgg gctcagggca gggcggttga gtttcccac 1440  
 ccatgtgggg acccttgggt agtcgcttga ttgggtagcc ctgaggaggc cgagatgcga 1500  
 tgggcccagg gccgtttcca aacacagagt caggcacgtg gaaggccag gaatcccctt 1560  
 20 ccctcagaggc aggagtgagg gaaaggagag ctgggccccg atttcacggc agccaggctg 1620  
 cagtgggcca ggtgtggtg gtccacgtgg cgtcgggggc ggggtctgat tcaaatcccg 1680  
 tggggctcgg ccttctcggc ccgtgctggc cgcgcctcca cacgggcttg ggggtggaccg 1740  
 ccgacactct agcaggtggc tatttctccc ttggaagag agccctcac ccatgctagg 1800  
 tgtttccctc ctgggtcagg agcgtggccg tgtggcaacc ccgggacctt aggtttattt 1860  
 atttgtttaa aaacattctg ggcctggctt ccgtgtgtgc taaatgggga aaagacatcc 1920  
 25 cacctcagca gagtactga gaggctgaaa ccgggtgtgt ggcttgactg gtgtgacttc 1980  
 aggtcattcc agaagtggct cagggaagtca gtgagaccag gtacatgggg ggctcaggca 2040  
 gtgggtgaga tgaggtacac ggggggctca ggcagtggtt gaggccaggc acatgggggg 2100  
 ctaggcact ggttgagatg aggtacacgg ggggctcagg cagagggtca gaccagggtac 2160  
 30 acgggggctc tgatcacacg cacatatgag cacatgtgca catgtgctgt ttcaggttag 2220  
 ccaggctctg gcacacctgc cccaaagtcc cagggaagctg agaggccaaa gatggaggct 2280  
 gacagggtg gcgcggtggc tcacacctgt agtcccagca ctttgggagg ccgaggcgag 2340  
 aggatccctt gagcccagga gtttaagacc agcctgagca acatagtaga accccatctc 2400  
 tatgaaaaat aaaaacaaaa attagctgaa catggtggtg tgcgcttga gttccaatc 2460  
 35 ttgggaggct gaagtgggag gatcactga gcccaggagg tggaaagctg agtgagctga 2520  
 gattgcacca ctgtactgca gcctgggtga cagagtgaga gccatctca acaacaacaa 2580  
 agaagactga caaatgcagt ttcttggaaa gaaacattta gtaggaaactt aacctacaca 2640  
 cagaagccaa gtcggtgtct cgggtgtcagt gagatgagat gatgggtcct cacaccatca 2700  
 cccagaccc aggttttatg caccacaggg gcgggtggct cagaagggat gcgcaggacg 2760  
 40 ttgatatacg atgacatcaa ggttgtctga cgaagggcag gattcatgat aagtaacctg 2820  
 tggatacaaa ggaacaatgg ataaactgga aaccttagag gccttcccgg aacaggggct 2880  
 aatcagaagc cagcatgggg ggctggcctc caggatggag ctgcttcagc ctccacatgc 2940  
 gtgttcatac agatgggtga cagaaacgca gtgtacctgt gcacacacag acacgcagct 3000  
 actgcacac acaagcacac acacagacat gcatgcacgc atccgtgtgt gtgcacctgt 3060  
 45 gcccagagg aaacctatgc atgtgcattc atgcacgcac acaggcaccg gtggggccat 3120  
 gcccacacc acgagcaccg tctgattagg aggcctttcc tctgacgctg tccgccatcc 3186  
 tctcag

<210> 18  
 <211> 781  
 50 <212> DNA  
 <213> Homo sapiens

<400> 18  
 gtatgtgcag gtgcctggcc tcagtggcag cagtgcctgc ctgctgggtg tagtgtgtca 60  
 55 ggagactgag tgaatctggg ctttaggaagt tcttaccctt tttcgcatca ggaagtgtt 120  
 taaccacaac actgtcaggc tcgtctgccc gccctctcgt ggggtgagca gacacactga 180  
 tggaaaggac aggagctgtc tgggagctgc catccttccc accttgctct gcctggggaa 240  
 gcgctggggg gcctgggtct tcctgtttgc cccatggtgg gatttggggg gcctggcctc 300  
 60 tcctgtttgc cctgtggtgg gattgggctg tctcccgtcc atggcactta gggcccttgt 360  
 gcaaacccag gccaaaggct taggaggagg ccaggccag gctacccac cctctcagg 420  
 agcagaggcc gcgtatcac acgacagagc cccgcgcgt cctctgtctt ccagtcaccg 480  
 tcctctgccc ctggacactt tgtccagcat caggagggtt tctgatccgt ctgaaattca 540  
 agccatgtcg aacctgcggt cctgagctta acagcttcta ctttctgttc tttctgtgtt 600

gtggaatatt cacctggaga agccgaagaa aacattttctg tctgtactcc tgcgggtgctt 660  
gggtcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagctttccc 720  
gtgtctcctg ggaggggagc tgggtctggc ctgtgactcc tcagcctctg ttttccccc 780  
g 781

5  
<210> 19  
<211> 536  
<212> DNA  
<213> Homo sapiens

10  
<400> 19  
gcaagtgtgg gtggaggcca gtgcggggccc cacctgccca ggggtcatcc ttgaacgccc 60  
tgtgtggggc gagcagcctc agatgtgtgt gaagtgcaga cggccccggg cctgaccctg 120  
ggggcctgga gccacgctgg cagccctatg tgattaaacg ctggtgtccc caggccacgg 180  
15 agcctggcag ggtccccaac ttcttgaacc cctgcttccc atctcagggg cgtatggctcc 240  
ccacgcttgg gagccttctg acccctgacc tgtgtcctct cacagcctct tccctggctg 300  
ctgccctgag ctctctgggt cctgagcaag ttctctcccc gccccgcccgc tccagcgtca 360  
ctgggtctgct tgtctgctcg ccccggtgga ggggtgtctg tcccttctact gaggttccca 420  
ccagccaggg ccacgaggtg cagggcctgc ctgcccgcc acccacacgt cctaggaggg 480  
20 ttggaggatg ccacctctgg cctcttctgg aacggagtct gatatttgcc ccgcag 536

<210> 20  
<211> 3179  
<212> DNA  
25 <213> Homo sapiens

<400> 20  
atctcatgtt tgaatcctaa tgtgcactgc atagacacca ctgtatgcaa ttacagaagc 60  
ctgtgagtga acggggtggg ggtcagtgcg ggcccatggc ctggctgtgc atttacggaa 120  
30 gtctatgagt gaattggggt gtggtcagtg cgggcccatt gcctggctgg gcctgggagg 180  
tttctgatgc tgtgaggcag gaggggaagg agggtagggg atagacagtg ggagccccc 240  
ccctggaaga cataacagta agtccaggcc cgaaggggcag caggatgctt gggggcccag 300  
cttgggcggc ggggatgatg gagggcctgg ccagggtggc agggatgatg ggggcccag 360  
35 ctggggtggc aggggtgatg gggggggctg gtctgggtgg cggggaagat ggggaagcct 420  
ggctggggccc cctcctcccc tgcctccccc ctgcagccgt ggatccggat gtgcttccct 480  
gggtgcacatc ctctgggcca tcagctttca tggagggtgg gggcaggggc atgacacatc 540  
cctgtataaa atccaggatt cctcctcctg aacgcccaca ctcagggtga aagtcacatt 600  
ccgctctctg ccattctctt aagagtagac caggattctg atctctgaag ggtgggtagg 660  
gtggggcagt ggggggtgtg gacacaggag gcttcagggt ggggctggtg atgctctctc 720  
40 atcctcttat catctcccag tctcatctct catcctctta tcatctccca gtctcatctg 780  
tcttctctct atctccagc tctcatctgt atcctcttac catctcccag tctcatctct 840  
tatcctctta tctcctagtc tcatccagac ttacctccca gggcgggtgc caggctcgca 900  
gtggagcttg acatacgtcc ttctcaggc agaaggaaat ggaaggattg cagagaacag 960  
gagggggcgg tcagaggggc gcagctcttg ggtgaagaaa cagcccccct tcagaagtgt 1020  
45 gcttgggcca cacgaaaccg agggccctgc gtgagtggct ccagagcctt ccagcagggt 1080  
cctgggtggg ccttatggta tggccgggtc ctactgagtg cactctggac agggcttctg 1140  
gtttgagtgc agcccggacg tgcctgggtg cggggtgggg gcttatggcc actggatatg 1200  
gcgtcattta ttgctgctgc ttcagagaat gtctgagtga ccgagcctaa tgtgtatgg 1260  
50 gggcccaagt ccacagactg tgcgtaaat gcactctggt gcctggagcc cccgtatagg 1320  
agctgtgagg aaggaggggc tcttggcagc cggcctgggg gcgcctttgc cctgcaaat 1380  
ggaaggagag gggcccgggc gccgtgggag gacgacctca agtgagagg tggacagaa 1440  
agggcgggga cttcccagga gcagaggccg ctgctcaggc acacctgggt ttgaatcaca 1500  
gaccaacagg tcaggccatt gttcagctat ccatcttcta caaagctcca gattcctgtt 1560  
tctccgggtg ttttttgggt aaattttact caggattact tatatttttt gctaaagtat 1620  
55 tagaccctta aaaaaggtat ttgctttgat atggcttaac tactaagca cctactttat 1680  
ttgtctgttt ttatttatta ttattattat tattagagat ggtgtctact ctgtcaccca 1740  
ggttgttagt gcagtggcac agtcatggct cgtgtagcc gcaaaccctc aggtcgaagt 1800  
gatcctccgg cctcagcttc ccagagtgtc gggattacag gtgtgagcca ctgcccttgc 1860  
ctggcacttt taaaaaccac tatgtaaggt cagggtccagt ggcttccaca cctgtcatcc 1920  
60 cagtgtttt ggaagccgag gcagaaggat tgtctgagtc caggagtgtg agaccagcat 1980  
gggtaacata gggagacccc atctctacaa aaaaatgcaa aagttatccg ggcgtgggg 2040  
ccagcatctg tagtcccagc tgctcgggag gctgagtggt aggtatcgctt gagcccggga 2100  
gggtatggct gcagtgagct gtgattgtac catcgactc cagcctgggc aacagagtga 2160  
gacctgtct caaaaaaaaa aaaaaaaaaa gaaggagaag gagaagagaa gaagaaggaa 2220  
65 gaaggaaaga gaagaagaag gaagaaggaa gaaagaaggaa gaaggaggcc tgctagggtc 2280

taggtagact gtcaaatctc agagcaaat gaaaataaca aagttttaa gggaaagaaa 2340  
aaccacagct ctttggactt ccttaggcct gaacttcac tcaagcagct tccttccaca 2400  
gacaagcgtg tatggagcga gtgagttcaa agcagaaagg gaggagaagc aggcaagggt 2460  
5 ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtggttgttt tcctgcctca 2520  
gccccacgct cctgccggtc ctgcacctgc tgtaaccgtc gatgttggtg ccagggtgcc 2580  
acctgggaag gatgctgtgc agggggcttg ccaaactttg gtgggtttca gaagccccag 2640  
gcacttgttg caggcacaaat tacagcccct ccccaaagat gccacgtcc ttctcctgga 2700  
acctgtgaat gtgtcacccg caaggcagag gctggtgaag gctgcagggtg gaatcacggc 2760  
10 tgccagtcag ccgatcttaa ggtcatcctg gattatctgg tgggcctgat atggccacaa 2820  
gggtccctag aagtgaagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880  
cactggccac tgctggcttt gagatggagg agggggtccc cagccaagga atgggggcag 2940  
ccgctccatg ctggaaaagc aagcaatcct ccccggtcct gagggcacac ggccctgccc 3000  
acgcctcgat ttcaggccag tgggacctgt ttcagctttc cggcctccag agctgtaaga 3060  
15 tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120  
cagtacagg aaatgaatac agggacagtt ctacagtgta ctctcagccc acccctggg 3179